D4

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68

A1

(11) International Publication Number: **WO 98/31835**

(43) International Publication Date: 23 July 1998 (23.07.98)

(21) International Application Number: PCT/US97/22974

(22) International Filing Date: 22 December 1997 (22.12.97)

(30) Priority Data:

60/035,382	16 January 1997 (16.01.97)	US
60/048,309	27 May 1997 (27.05.97)	US

(71) Applicants *(for all designated States except US)*: UNIVERSITY OF MARYLAND COLLEGE PARK [US/US]; Office of Technology Liaison, 4312 Knox Road, College Park, MD 20742 (US). UNIVERSITY OF PITTSBURGH [US/US]; Office of Technology Management, Room 9110, William Pitt Union, Pittsburgh, PA 15260 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: HAGBERG, James, M. [US/US]; 10941 Hilltop Lane, Columbia, MD 21044 (US). FERRELL, Robert, E. [US/US]; 206 Maple Avenue, Pittsburgh, PA 15218 (US).

(74) Agent: POULOS, James, A., III; Nikaido, Marmelstein, Murray & Oram LLP, Metropolitan Square, Suite 330, 655 Fifteenth Street, Washington, DC 20005-5701 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: GENETIC METHODS FOR IDENTIFYING INDIVIDUALS FOR IMPROVING WELL BEING AND PERFORMANCE THROUGH EXERCISE

(57) Abstract

Methods for improving the health and/or performance of a person having identified genotypes through exercise.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

GENETIC METHODS FOR IDENTIFYING INDIVIDUALS FOR IMPROVING WELL BEING AND PERFORMANCE THROUGH EXERCISE

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a non-provisional application claiming benefit of provisional application No. 60/035,382, filed January 16, 1997 and 60/048,309, filed May 27, 1997, both applications herein incorporated by reference.

FIELD OF THE INVENTION

This invention relates to identifying alleles associated with phenotypes and recommending lifestyle changes based on the allele identification.

BACKGROUND OF THE INVENTION

Studies have shown that individuals suffering from hypertension, insulin resistance, arthritis, unfavorable cholesterol profiles, and other ailments can alleviate their symptoms or improve their conditions through exercise.

Unfortunately, some individuals, no matter how rigorously they exercise or train, are unable to improve their conditions, and yet still others benefit to a much greater extent than predicted.

These results underscore the fact that many factors contribute to one's well-being. Such factors include, for instance, behaviors such as diet and exercise, genetic makeup, and environment.

While behavior can be controlled or altered, and our environments can be regulated, our genetic makeup, at least up until now, is predetermined and set at birth. By identifying

the genetic makeup of a population and recognizing that some individuals of the population will benefit from a change of behavior, and still others will benefit to a much greater extent, and still others do not benefit at all, one hypothesizes that the genotype of an individual is a predictor of a result achieved through a change of behavior.

Thus, an object of this invention is to identify individuals possessing a certain genotype and associated ailment, and to determine if the health of that individual can be improved by altering behavior.

Another object of the invention relates to identifying individuals having certain apolipoprotein (APO) E and lipoprotein (LPL) genotypes and recommending present or future changes in behavior that will improve the cholesterol profile of such person.

Another object of the invention relates to identifying patients having certain angiotensin converting enzyme (ACE), APO E and LPL alleles and recommending present or future behavioral changes that will improve hypertension in that individual.

Another object of the invention relates to identifying women having certain Vitamin D receptor (VDR) alleles and recommending present or future changes in behavior that will improve the bone mineral density of such women.

Another object of the invention relates to identifying persons having certain ACE alleles for the purpose of quantifying their current and future risk of developing cardiovascular disease.

DETAILED DESCRIPTION OF THE INVENTION

For purposes of this invention a lifestyle change may involve a course of action to overcome or enhance a particular genotype. The course of action may be, for example, a change in diet, starting an exercise program or both, a change in environment, or for instance, changing an exercise regime, etc.

Cholesterol and APO E

Apolipoprotein E (APO E)[SEQ. ID NO.2] plays a central role in cholesterol transport, total cholesterol and low-density lipoprotein cholesterol concentrations, and contributes to coronary heart disease risk. APO E is a ligand for lipoprotein receptors. Physiologically, its most important function is to mediate specific uptake of plasma very low-density lipoproteins, chylomicron remnants, and intermediate-density lipoprotein by the liver. The APO E gene disclosed in Hixson et al, "Restriction isotyping of human apolipoprotein E by gene amplification and cleavage with HhaI." Journal of Lipid Research Volume 31, 1990 (herein incorporated by reference) [SEQ. ID NO. 1], is polymorphic with three common alleles designated E2, E3, and E4, resulting in six major Apo E genotypes: E2/2, E3/2, E4/2, E3/3, E4/3 and E4/4. The Apo E3 allele is the most common allele in populations studied. Isoform E3 is distinguished by cysteine at position 112 (112 cys) and arginine at codon 158 in the receptor binding region of Epo E. Codon 158 is a positively charged

region of the molecule that binds to low-density lipoprotein receptors. The E4 isoform (112arg and 158arg) is associated with increased levels of total cholesterol and betalipoprotein. The E4 allele has an amino acid substitution at Codon 112 that has arginine in place of cysteine, and this region appears to reduce disulfide bonding of Apo E with other free sulfhydryl-containing proteins. Consequently, APO E4 may be more readily transferred from high-density lipoprotein to chylomicron remnants than APO E3 resulting in enhanced receptor-mediated clearance of low-density lipoprotein, hepatic cholesterol accumulation, down-regulation of lipoprotein receptors, and subsequently raised concentration of serum cholesterol, giving rise to increased susceptibility to heart disease. Most patients with type III hyperlipidemia (a condition characterized by the accumulation in plasma of remnants of the metabolism of triacylglycerol-rich lipoproteins), are homozygous for the E2 isoform (112cys and 158cys) that binds with reduced affinity to cellular receptors. In population studies, the E2 isoform is associated with decreased levels of cholesterol and betalipoprotein.

Lipoproteins with Apo E2 bind less well to low-density lipoprotein receptors, so clearance is slower. Low-density lipoprotein receptors are subsequently up-regulated, and serum cholesterol concentrations are reduced in the majority of individuals with the E2 allele. However, 1-10 % of individuals with the E2/2 genotype develop type III hyperlipidemia. This variability in response suggests that

other genes and/or environmental factors play important roles in the development of overt type III hyperlipidemia in subjects with the E2/2 genotype.

The APO E polymorphism has been consistently shown to be associated with plasma concentrations of total and low-density lipoprotein cholesterol.

Identifying Individuals with Apo E Alleles.

In a first study, obese sedentary hypertensive men, 50-65 years of age, had DNA collected from them, and this DNA was analyzed for the presence of the Apo E alleles identified by the method below:

Typing of the subjects was determined by using an isoelectric focusing-immunoblotting method as described previously Kamboh et al, "Impact of apolipoprotein E polymorphism in determining interindividual variation in total cholesterol and low density lipoprotein cholesterol in Hispanics and non-Hispanics whites". Atherosclerosis 1993;98:201-11 (herein incorporated by reference). Other methods for identifying APO E alleles include using the polymerase chain reaction (PCR) see Kamboh et al, "The relationship of APO E polymorphism and cholesterol levels in normoglycemic and diabetic subjects in a biethnic population from the San Luis Valley, Colorado." Atherosclerosis 1195:112:145-149. DNA was extracted from lymphocytes as described by Miller et al, "A simple salting out procedure for extracting DNA from human nucleated cells." Nucleic Acids Research 1988:16:1215. Genomic DNAs (0.5-1.0 μg) are amplified by using a forward primer E1, 5'-GCGGACATGGAGGACGTG-

3' [SEQ. ID NO. 3](codons 106-111) and a reverse primer E2, 5'-GGCCTGGTACACTGCCAG-3' [SEQ. ID NO. 4] (codons 159-164). The 50-μL reaction mixture consisted of 5 μL 10 x reaction buffer (100 mmol Tris-HCl/L, pH 8.9, 500 mmol KCl/L, 15mmol $MgCl_2$/L, 3 μL dimethylsulfoxide, 0.75 μL (0.3 μmol) of each primer; 0.6 μL each of (10mmol/L) dATP, dGTP, dCTP, and dTTP; 0.3 μL *Thermus aquaticus* (*Taq*) DNA polymerase (Perkin Elmer Cetus Inc., Foster City, CA); and 37.8 μL sterile deionized water, which was added to the DNA template. Twenty-five microliters mineral oil was placed over the final PCR reaction mixture. After a denaturation step of 8 min at 95°C, amplification is achieved by 30 cycles of denaturation (1 min at 95°C), annealing (1 min at 57°C), and extension (2 min at 72°C), followed by extension for 5 min at 72°C. The nucleotide amplified product is digested directly with the restriction enzyme *Hha*I (New England Bio Labs, Beverly, MA). The digested DNAs are separated on 8% nondenaturing polyacrylamide gel in 1 x Tris borate EDTA buffer followed by staining with ethidium-bromide solution and the APO E polymorphism was typed by visualization under ultraviolet light.

These subjects were exposed with aerobic exercise training and were given the same diet. Specifically, to eliminate the effect of diet, all subjects were instructed in the principles of weight-maintaining American Heart Association (AHA) step I diet over an 8-week period before baseline testing. This diet consisted of 50-55% of calories as carbohydrate, 30-35 % as fat, 15-20% as protein, 300-350 mg/day of cholesterol, and 3 g/day of sodium. These subjects

were counseled weekly to maintain their diet consumption throughout the length of monitoring. Adherence was monitored by registered dieticians who reviewed weekly food records and body weights and calculated dietary consumption from biweekly 7-day food records. At baseline and after the intervention, the subjects were weight-stable for 4 weeks before testing. During this period, the subjects were instructed to maintain their body weight within 1 kilogram.

These subjects took part in an aerobic exercise program that met 3 times per week. Exercise training consisted of stationary cycling and walking and jogging on a treadmill starting at 50-60% of each individual's heart rate reserve for three five- to ten-minute periods. Target heart rate was calculated for each individual with the equation of Karvonen et al., "The effects of training heart rate: a longitudinal study". *Ann. Med. Exp. Biol. Fenn.* 35:307-315 (1957).

Training intensity was gradually increased by five to ten percent of the heart rate reserve every month. At three months, the maximal oxygen consumption (VO_{2max}) test was repeated, and the intensity was adjusted until forty minutes of training per session at an intensity of seventy-five to eighty-five percent of heart rate reserve was achieved. All training sessions were supervised by the research staff, and the subjects were instructed by a dietician to increase the caloric intake to offset the increase in energy expenditure due to the increased physical activity.

Serum total cholesterol and high-density lipoprotein cholesterol and high density lipoprotein 2 cholesterol levels

were measured for the subjects. Subjects possessing at least one APO E2 allele increased their high-density lipoprotein cholesterol (HDL-C) levels over seven times more with exercise training than those with only APO E3 or E4 alleles (Table 1). High-density lipoprotein cholesterol 2 (HDL_2-C) increases with exercise training were also dramatically larger in APO E2 vs. APO E3 and E4 individuals (Table 1).

Table 1: Plasma Lipoprotein-lipid changes with exercise training as a function of genotype

TABLE 1

	Change with Intervention		
	Cholesterol	HDL-C	HDL_2-C
APOE 2/2 and 2/3 genotypes (n=3)	-17.3 ± 13.9	12.1 ± 6.9	8.8 ± 6.6
APOE 3/3 and 3/4 genotypes (n=16)	-18.8 ± 3.8 P=NSD	1.6 ± 1.1 P=0.02	-0.1 ± 1.0 P=0.04
Values are mean ± SE. Probabilities are for two-tailed tests.			

"n"= the number of subjects in each group
"p" indicates statistical probability for the difference between group

In a second study, subjects with lipoprotein lipase PvuII-/- genotype also had substantially greater exercise training-induced increases in plasma HDLC and HDL_2C levels than those with +/+, or +/- genotypes (Table 2).

Lipoprotein lipase (LPL) [SEQ. ID NO. 6] is an enzyme that catalyzes the breakdown of triglycerides in the plasma to release free fatty acids. This hydrolysis also influences the metabolism of circulating lipoproteins. LPL has also been shown to enhance the triglyceride-rich chylomicron binding to low density lipoprotein receptor-related proteins. Thus, LPL may also be an

important regulator of chylomicron metabolism. The LPL gene [SEQ. ID NO. 5] is located on human chromosome 8p22. It is approximately 35 kilobases long and has 10 functionally differentiated exons. Two primary polymorphic variations occur at the LPL gene locus in frequencies that are important on a population basis; these two markers are detected by PvuII and HindIII. There are three genotypes at each site with the alleles for both denoted as "+" or "-" based on the presence or absence of a restriction site at the LPL locus with PvuII or HindIII. Thus, for both PvuII and HindIII there are three genotypes: +/+, +/-, and -/-. PvuII polymorphic variations have been reported to be associated with variations in plasma triglyceride levels. HindIII variations have previously been shown to be associated with hypertriglyceridema, coronary artery disease, plasma total and HDL-cholesterol levels, and plasma insulin levels.

One individual with the APO E2 and LPL PvuII -/- genotypes elicited the largest HDL-C and HDL_2-C increases of the nineteen men in the study. Training-induced changes in total cholesterol and triglyceride levels were similar in all APO E and LPL PvuII genotype groups. No other training-induced changes in variables that affect plasma lipoprotein-lipids (including dietary habits, VO_{2max}, body weight, and body composition) differed among either APO E or LPL PvuII genotype groups. Thus, these results show that APO E or LPL PvuII genotype identifies individuals who will improve their blood cholesterol profile with exercise training.

TABLE 2

	Change with Intervention		
	Cholesterol	HDL-C	HDL_2-C
LPL PvuII -/- genotype (n=3)	-14.0 ± 14.5	11.4 ± 7.3	8.5 ± 6.7
LPL PvuII +/- and +/+ genotypes (n=15)	-19.0 ± 4.0 P=NSD	1.8 ± 1.2 P=0.05	-0.2 ± 1.1 P=0.06
Values are mean ± SE. Probabilities are for two-tailed tests.			

"n"= the number of subjects in each group
"p" indicates statistical probability for the difference between group

In a third study, cross-sectional differences in plasma HDL-C and HDL_2-C levels were found that provide further evidence that gene markers identify individuals most likely to improve their blood cholesterol profile with exercise training (Table 3). Postmenopausal women were studied to assess the effect of physical activity behavior on cardiovascular disease risk factors.

Women were classified as postmenopausal by self-reported lack of menses for greater than two years and elevated levels of follicle stimulating hormone and luteinizing hormone. Women were classified as sedentary if they had not participated in regular aerobic activity for greater than two years. Women who participated in aerobic exercise for greater than 90 minutes/week, for greater than three years, but who were not training for endurance-based competitive events were classified as physically active.

Endurance trained women were defined as those undergoing rigorous exercise training more than 4-5 times per week for at

least two years in preparation for competitive endurance-based events (primarily long distance running).

It was found that the APO E genotype did not affect plasma lipoprotein lipids in either sedentary or physically active women. However, among endurance-trained women, those with at least one APO E2 allele had substantially higher HDL-C and HDL_2-C levels than those with only APO E3 or E4 alleles. Furthermore, endurance-trained women with the APO E2 genotype had higher HDL-C and HDL_2-C levels than the sedentary or physically active women, whereas HDL-C and HDL_2-C levels were the same in all APO E3 or E4 women regardless of physical activity habits. No other differences were evident between the APO E genotype groups of endurance-trained women that could affect plasma lipoprotein lipids including dietary habits, VO_{2max}, body weight, body composition, regional distribution of fat, running mileage, years of running, years postmenopausal, and hormone replacement status. Thus, these data also show that genetic markers identify individuals most likely to improve their blood cholesterol levels with exercise training.

Table 3: HDL-C and HDL_2-C levels as a function of genotype and exercise training status

	Those with APOE 2/2 or 2/3 genotype		Those with all other APOE genotypes	
	HDL-C	HDL_2-C	HDL-C	HDL_2-C
Sedentary + Active Women	57	32	58	33
	(n=9)		(n=27)	
Endurance-trained Women	75	55	62	36
	(n=4)		(n=18)	

The values reported are mean values for each group
n= the number of subjects in a group

The difference between the two genotype groups within the endurance-trained women is p=0.20 for HDL-C and p=0.035 for HDL_2-C. The difference between endurance-trained women and the combined group of sedentary and physically-active women within the APO E 2/2 and 2/3 genotype is p=0.15 for HDL-C and p=0.08 for HDL_2-C.

Thus, high levels of physical activity interact with a woman's genotype such that only well-trained postmenopausal women with the APO E 2/2 or 2/3 genotype have better HDL-C and HDL_2-C levels compared to their sedentary or physically active peers.

With this knowledge, clinical diagnostic kits made available can be used in identifying individuals most likely to improve their blood cholesterol profiles with exercise training.

Hypertension and ACE

ACE

Angiotensin converting enzyme (ACE) [SEQ. ID NO. 8] is the enzyme responsible for catalyzing the conversion of angiotensin I, a relatively inactive tissue and plasma vasopressor hormone, into the potent and highly active vasopressor hormone, angiotensin II. This cascade of reactions is part of the renin-angiotensin-aldosterone system that has long been known to be an important regulator of arteriolar relaxation and vasoconstriction, and hence blood pressure, in humans and animals. The ACE gene [SEQ. ID NO. 7] is polymorphic with two common alleles designated I and D, resulting in three genotypes: II, ID, and DD. The D allele has a 287-base pair marker in intron 16 of the ACE gene deleted, whereas the I allele has the 287-base pair marker inserted. The D allele is associated with increased levels of ACE in both plasma and ventricular tissues. Increased levels of ACE will clearly contribute to increased myocardial and vascular smooth muscle growth and increased arteriolar vasoconstriction. Thus, the presence of the D allele would by hypothesized to have deleterious effects on the cardiovascular system and, in fact, the D allele has been associated with increased risk of left ventricular hypertrophy, cardiovascular disease, and sudden cardiovascular death. The D allele was also originally believed to be more prevalent in hypertensives than normotensives; however, this relationship has generally not been substantiated in more recent studies.

Identifying Individuals with Different ACE Alleles to Predict Blood Pressure Changes with exercise Training

Obese sedentary hypertension men, 50-65 years of age had DNA collected from them and the DNA was analyzed for the presences of ACE alleles. Typing of these individuals was conducted by isolating high molecular weight genomic DNA from whole blood mononuclear cells by the procedure of MIller et al (1988). ACE genotyping was carried out by polymerase chain reaction amplification using the forward primer 5'-CCGTTTGTGCAGGGCCTGGCTCTCT-3' [SEQ. ID No. 9] and reverse primer 5'-CAGGGTGCTGTCCACACTGGACCCC-3' [SEQ. ID NO.10] and the following cycling conditions: denaturation at 95°C for 5 minutes followed by thirty cycles of 30 sec. Denaturation at 94°C, 15 sec annealing at 58°C, 30 sec. extension at 72°C. Amplimers were resolved on 2% agarose gels and genotypes assigned by direct comparison to samples of known genotype. The I (insertion) allele yielded a band of 490 bp and the D (deletion) allele a band of 190 bp. Heterozygotes were typed by the presence of both bands plus a heteroduplex band migrating at an approximately 370 bp. (Tiret et al. "Evidence from combined segregation and linkage analysis that a variant of the angiotensin-1 converting enzyme (ACE) gene controls plasma ACE levels" AM J. Hum Genet 1992; 51:197-205) reported above.

These subjects were exposed with aerobic exercise training and were given the same diet. All subjects were instructed in the principles of weight-maintaining American Heart Association step I diet over an 8-week period before baseline testing as

reported at page 6. These subjects were counseled weekly to maintain their diet as reported at page 6. Adherence was monitored as reported at page 6.

These subjects took part in an aerobic exercise program that met 3 times per week as reported at pages 7.

Subjects had their blood pressure measured weekly for 4 weeks prior to and following the completion of the exercise training intervention. Blood pressure measurements were made using a mercury sphygmomanometer and stethoscope according to standards established by the American Heart Association for cuff size, Korotkoff sounds, pre-measurement rest, and the number, timing, and averaging of the blood pressure results. The final values at baseline and after the intervention represent the average of 12 independent blood pressure measurements (3 on each of the 4 measurement days).

Results of this study indicate that ACE genotype identifies hypertensive individuals that reduce their systolic and diastolic blood pressure with exercise training. Those subjects with at least one insertion ("I") ACE allele decreased their diastolic blood pressure with exercise training approximately 7 times more than those with only deletion ("D") ACE alleles (Table 4). Subjects with at least one insertion ACE ("I") allele decreased their systolic blood pressure with exercise training over twice as much as those with only deletion ACE ("D") alleles (Table 4). Subjects' baseline characteristics prior to exercise training did not differ among ACE genotype groups. Furthermore, no other training-induced changes in variables that affect systolic and diastolic blood pressure, including dietary habits, VO_2max, body

weight, and body composition, differed among ACE genotype groups. Thus, these results show that ACE genotype is a strong independent indicator of those individuals who will reduce their systolic and diastolic blood pressure with exercise training.

Table 4: Change in Blood Pressure with Exercise Training in Hypertensives as a Function of ACE Genotype

ACE genotype	Change in BP with Exercise Training	
	Systolic BP	Diastolic BP
II and ID genotype (n=11)	-9.8 ± 9.6	-10.0 ± 6.0
DD genotype (n=8)	-4.7 ± 8.0 (P=0.14)	-1.4 ± 5.0 (P<0.005)
Values are mean ± SD. Probabilities are for two-tailed t-tests.		

In yet another method, applicants have discovered that individuals with different LPL PvuII genotypes, after exercise training will exhibit different results relative to systolic and diastolic blood pressure not predicted from initial screenings.

Specifically, the same obese sedentary hypertensive men age 50-65 earlier described had DNA collected from them, and this DNA was analyzed for the presence of genetic variations at two critical restriction sites at the lipoprotein lipase (LPL) gene locus. (See page 9 supra)

DNA samples were subjected to amplification by the polymerase chain reaction in a Perkin-Elmer Cetus DNA Themal Cycler. One set of primers was derived from sequences between exons 8 and 9 in the, LPL gene to amplify the sequence around a HindIII restriction site in intron 8 (the forward primer was 5'-

TTTA GGCCTGAAGTTTCCAC-3' [SEQ ID NO. 11] and the reverse primer was 5'CTCCCTAGAAGAGAAGATC-3' [SEQ ID NO. 12] as described (Kirchgessner TG. et al., "Organization of human lipoprotein lipase gene and evolution of the lipase gene family." Proc National Academy of Science 86: 9647-9651, 1989). The amplified fragment had a size of 1.3 kilobases. The second set of primers was from the DNA sequences flanking the PvuII reaction site. In intron 6 (the forward primer was 5'TAGGAGGTTGAGGCACCTGTGC-3' [SEQ. ID NO. 13] and the reverse primer was 5'GTGGGTGAATCACCTGAGGTC-3' [SEQ. ID NO. 14] as described (Oka K et al., "Nucleotide sequence of PvuII polymorphic site at the human lipoprotein lipase gene locus." Nucleic Acid Research 17: 6752, 1989). This amplified fragment was 858 base pairs long.

The 50 ul reaction mixture contained 1 x PCR buffer (10mM Tris, pH 8.3. 50mM KCl, 1.5mM $MgCl_2$,), dNTPs at 200uM, 0.3uM each primer, 0.5 ug genomic DNA, and 1.25 units of Taq DNA polymerase. Amplification of the region flanking the HindII site was carried out for 33 cycles at 95°C for 1 min, at 60°C for 2 min, and at 72° for 2 min. For amplification around the PvuII site, the conditions were the same except for annealing at 70°C and 25 cycles. Amplified products were digested with HindIII or PvuII and the resulting fragments separated on 2% agarose gels. After digestion with, HindIII the presence of the restriction site (+ allele) resulted In fragments of 600 and 700 base pairs. The Presence of the PvuII site (+ allele) yielded fragments of 266 and 592 base pairs.

These subjects were exposed with aerobic exercise training and were given the same diet as reported at page 6. Specifically,

to eliminate the effect of diet, all subjects were instructed in the principles of weight-maintaining American Heart Association step I diet over an 8-week period before baseline testing as reported at page 6. These subjects were counseled weekly to maintain their diet consumption throughout the length of monitoring. Adherence was monitored as reported at page 10.

These subjects took part in an aerobic exercise program as reported at pages 7.

Prior to having undergone exercise training and after undergoing the exercise regimen and associated diet (the same regimen and diet as reported at page 6 above), subjects were tested for systolic and diastolic blood pressure levels. Initial levels did not differ between hypertensive individuals with different LPL PvuII genotypes. However, the changes in blood pressure resulting from exercise training did differ among LPL PvuII genotype groups (see Table 5). Those having only + alleles decreased both their systolic and diastolic blood pressures more with exercise training than those with the -/- or the +/- genotype. In fact, both genotype groups decreased their systolic and diastolic blood pressure significantly with exercise training. However, those men with the +/+ LPL PvuII genotype tended to decrease both their systolic and diastolic blood pressure more than men with the +/- or -/- genotype. The subjects baseline characteristics prior to exercise training did not differ among LPL PvuII genotype groups, except that the -/- and the +/- genotype men were somewhat older that the +/+ men (63.8 vs 56.5 years). Furthermore, no other training-induced changes in variables that affect systolic and diastolic blood

pressure, including dietary habits, VO_2max, body weight, and body composition, differed among LPL PvuII genotype groups.

TABLE 5

Changes in blood pressure with exercise training in the two LPL PvuII genotype groups

Variable	-/- and +/- n=14	+/+ n=4
Systolic blood pressure, mmHg		
-initial	155 ± 10	156 ± 13
-change with training	-6 ± 8	-14 ± 11
Diastolic blood pressure, mmHg		
-initial	95 ± 7	97 ± 8
-change with training	-5 ± 6	-9 ± 9

"n"= the number of subjects in each group

Thus, these results indicate that LPL PvuII genotype is an indicator of those individuals most likely to reduce their systolic and diastolic blood pressure the most with exercise training.

In these same individuals, initial systolic blood pressure levels did not differ between hypertensive individuals with different LPL HindIII genotypes, but initial diastolic blood pressures were somewhat higher in LPL HindIII +/+ and +/- genotype compared to -/- genotype men (94 vs. 86 mmHg). However, the changes in blood pressure resulting from exercise training did differ among LPL HindIII genotype groups (see Table 6.) Those having at least one + allele decreased both their systolic and diastolic blood pressures significantly with exercise training. Furthermore, men with the +/+ or +/- LPL HindIII genotype decreased both their systolic and diastolic blood pressures more than men with the -/- genotype. The subjects'

other baseline characteristics prior to exercise training did not differ among LPL HindIII genotype groups. Furthermore, no other training-induced changes in variables that affect systolic and diastolic blood pressure, including dietary habits, VO_2max, body weight, and body composition, differed among LPL HindIII genotype groups.

TABLE 6
Changes in blood pressure with exercise training in the two LPL Hind III genotype groups

Variable	+/+ and +/- n=15	-/- n=3
Systolic blood pressure, mmHg -initial -change with training	 155 ± 11 -6 ± 8	 149 ± 5 +3 ± 4
Diastolic blood pressure, mmHg -initial -change with training	 94 ± 6 -9 ± 6	 86 ± 6 2 ± 3

"n"= the number of subjects in each group

Thus, these results indicate that LPL HindIII genotype is an indicator of those individuals most likely to reduce their systolic and diastolic blood pressure the most with exercise training.

Vitamin D receptor gene [SEQ. ID NO. 15] and Bone density

The vitamin D receptor [SEQ. ID NO. 16] plays a central role in the regulation of calcium metabolism, hence it also has a critical role in determining bone homeostasis. The Vitamin D receptor combines with other nuclear factors to activate the transcription of a large number of target genes including osteocalcin. The vitamin D receptor has been found to have a polymorphic variation that is detected with BsmI. Alleles with the presence of the BsmI restriction site are denoted as "b" and

those without the BsmI restriction site are denoted as "B" giving rise to three distinct vitamin D receptor genotypes: bb, Bb, and BB. These genotypes have previously been found to be related to circulating osteocalcin levels and bone mineral density in both young and postmenopausal women. Furthermore, these genotypes have been found to be related to risk of hip fracture and the rate of bone loss after the menopause in women.

In order to identify these alleles, high molecular weight DNA was extracted from peripheral leukocytes by standard methods.(SA Miller et al. A simple salting out procedure for extracting, DNA from human nucleated cells. Nucleic Acids Research 16: 1215-1218, 1988). DNA was amplified by PCR, using amplification primers as described by NA Morrison, et al. "Contribution of trans-acting factor alleles to normal physiologic variability, vitamin D receptor gene polymorphism and circulating osteocalcin." Proc National Acad Sci 89: 6665-6669, 1992). Each PCR was performed using 60 ul final reaction volume containing 110-200 ng DNA, 0.46 uM of each primer, 185 uM of dNTP mixture, 50mM KCI, 10 mM Tris-HCI pH 9.0, 1.5 mM MgCl, 0.1% Triton X-100, 0.8 units Taq polymerase). After amplification, 5 units of BsmI (New England Biolabs, Beverly, MA) was added to 16 uL of amplified product for digestion at 65° C. Each digested sample was loaded onto 2% agarose gels containing ethidium bromide and electrophoresed for 3 hours at 90 volts, After electrophoresis, the DNA fragments were visualized by UV illumination and fragment sizes were estimated by comparison to a 1 kilobase size ladder run on the same gel. The presence of a polymorphic restriction site was specified as "b", whereas the

absence of this site was specified as "B".

Fifty-one healthy postmenopausal women were recruited into six specific groups based on their hormone replacement therapy status and habitual levels of activity. Women were classified as postmenopausal by self-reported lack of menses for greater than two years and elevated levels of follicle stimulating hormone and luteinizing hormone. Women were classified as sedentary if they had not participated in regular aerobic activity for greater than two years. Women who participated in aerobic exercise for greater than 90 minutes/week, for greater that three years, but who were not training for endurance-based competitive events were classified as physically active. Endurance trained women were defined as those undergoing rigorous exercise training more than 4-5 times per week for at least two years in preparation for competitive endurance-based events (primarily long distance running).

Hip and lumbar spine bone mineral density were determined on all women in the morning after an overnight fast while wearing standard hospital gowns using a Lunar Corporation DPX-L dual energy X-ray absorptiometry system. Scans were obtained and analyzed using conventional methods and software. Quality control measures were recorded prior to each scan to ensure validity of the bone mineral density result.

The results of testing for the bone density of the women are reported below:

TABLE 7

TROCHANTER BONE MINERAL DENSITY

Habitual Physical Activity Level	Vitamin D Receptor Genotype BB	Bb	bb
Sedentary	0.704 (n=4)	0.687 (n=5)	0.628 (n=9)
Physically Active	0.789 (n=5)	0.722 (n=13)	0.737 (n=2)
Endurance Athletes	0.691 (n=6)	0.695 (n=10)	0.696 (n=7)

LUMBAR VERTEBRAE TWO-FOUR BONE MINERAL DENSITY

Habitual Physical Activity Level	Vitamin D Receptor Genotype BB	Bb	bb
Sedentary	0.915 (n=4)	0.812 (n=5)	0.897 (n=9)
Physically Active	1.064 (n=5)	0.997 (n=13)	0.904 (n=2)
Endurance Athletes	0.882 (n=6)	0.899 (n=10)	0.905 (n=7)

Values are means expressed in g/cm^2

Hormone replacement therapy had no effect on bone mineral density in these women. These results show that low to moderate levels of habitual physical activity interact with the BB vitamin D receptor resulting in increased bone mineral density of postmenopausal women.

The ACE locus and VO_2max

In yet another study healthy postmenopausal women 50-75 years of age underwent VO_2max testing on a treadmill to determine the maximal amount of oxygen they could consume (VO_2max) when exercising to their own individual maximal capacity. The women were almost equally divided between women on and not on hormone

replacement therapy. In addition, the women were approximately equally divided between groups that were sedentary, physically-active, and endurance trained. Hormone replacement therapy did not affect VO_2max. Habitual physical activity level affected VO_2max with the sedentary women having the lowest, the physically active having intermediate, and the endurance trained women having the highest VO_2max values. ACE genotype also affected VO_2max in these women after accounting for the effects of habitual physical activity level on VO_2max see Table 8 below:

TABLE 8

VO_2max values in postmenopausal women with different habitual physical activity levels as a function of ACE genotype.

ACE Genotype	VO_2max
II (n = 10)	2.33± 4.18
ID (n = 25)	-0.02 ± 4.25
DD (n = 15)	-1.52 ± 2.56

Values in Table 8 are expressed as mean ± standard deviation in units ml/kg/min. Values in parentheses indicate the number of subjects in each group. To account for the different physical activity levels of the women, which independently affects VO_2max, each woman's VO_2max was expressed as a difference from the average value for their respective habitual physical activity level group (sedentary, physically-active, endurance-trained). Thus, positive values indicate results above the average based on physical activity level and negative values indicate results below the average based on physical activity level.

Both the II and DD genotype groups were significantly

different from zero (P=0.04 and 0.02 respectively), whereas the ID genotype group was clearly not significantly different from zero. The difference between II and DD genotype groups was also statistically significant (P=0.02). The differences between the two homozygote groups when compared to the heterozygote ID genotype group both approached statistically significant (p=0.14-0.15).

These results show that a specific gene marker can identify an individual having a higher or lower VO_2max value compared to their peers. Thus, an individual with an ACE II genotype can be directed and would be expected to have the ability to compete with great success in activities such as running, cycling and swimming. Show and performance animals such as steeplechase horses and thoroughbreds, and greyhounds etc. having a similar genotype would also be expected to have the ability to compete with great success.

Because a low VO_2max is a risk factor for cardiovascular disease, ACE genotype will also help to identify individuals at lower and greater risk for cardiovascular disease.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: HAGBERG, JAMES M
FERRELL, ROBERT E

(ii) TITLE OF INVENTION: GENETIC METHODS FOR IDENTIFYING INDIVIDUALS FOR IMPROVING WELL BEING OR PERFORMANCE THROUGH EXERCISE

(iii) NUMBER OF SEQUENCES: 16

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: WATSON COLE STEVENS DAVIS,P.L.L.C.
(B) STREET: 1400 K STREET NW
(C) CITY: WASHINGTON
(D) STATE: D.C.
(E) COUNTRY: USA
(F) ZIP: 20005-2477

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: POULOS III, JAMES A
(B) REGISTRATION NUMBER: 31,714
(C) REFERENCE/DOCKET NUMBER: JAP70494

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 202 628 0088
(B) TELEFAX: 202 628-8034

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 244 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 9..233

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
TAAGCTTG GCA CGG CTG TCC AAG GAG CTG CAG GCG GCG CAG GCC CGG CTG        50
          Ala Arg Leu Ser Lys Glu Leu Gln Ala Ala Gln Ala Arg Leu
            1               5                  10

GGC GCG GAC ATG GAG GAC GTG CGC GGC CGC CTG GTG CAG TAC CGC GGC        98
Gly Ala Asp Met Glu Asp Val Arg Gly Arg Leu Val Gln Tyr Arg Gly
 15                  20                  25                  30

GAG GTG CAG GCC ATG CTC GGC CAG AGC ACC GAG GAG CTG CGG GTG CGC       146
Glu Val Gln Ala Met Leu Gly Gln Ser Thr Glu Glu Leu Arg Val Arg
                35                  40                  45

CTC GCC TCC CAC CTG CGC AAG CTG CGT AAG CGG CTC CTC CGC GAT GCC       194
Leu Ala Ser His Leu Arg Lys Leu Arg Lys Arg Leu Leu Arg Asp Ala
            50                  55                  60

GAT GAC CTG CAG AAG CGC CTG GCA GTG TAC CAG GCC GGG GCGAATTCTG        243
Asp Asp Leu Gln Lys Arg Leu Ala Val Tyr Gln Ala Gly
        65                  70                  75

T                                                                     244
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 75 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Ala Arg Leu Ser Lys Glu Leu Gln Ala Ala Gln Ala Arg Leu Gly Ala
  1               5                  10                  15

Asp Met Glu Asp Val Arg Gly Arg Leu Val Gln Tyr Arg Gly Glu Val
            20                  25                  30

Gln Ala Met Leu Gly Gln Ser Thr Glu Glu Leu Arg Val Arg Leu Ala
        35                  40                  45
```

Ser His Leu Arg Lys Leu Arg Lys Arg Leu Leu Arg Asp Ala Asp Asp
50 55 60

Leu Gln Lys Arg Leu Ala Val Tyr Gln Ala Gly
65 70 75

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GCGGACATGG AGGACGTG 18

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GGCCTGGTAC ACTGCCAG 18

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3549 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens
(F) TISSUE TYPE: Adipose tissue

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 175..255

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 175..1602

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION: 256..1599

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
CCCCTCTTCC TCCTCCTCAA GGGAAAGCTG CCCACTTCTA GCTGCCCTGC CATCCCCTTT      60

AAAGGGCGAC TTGCTCAGCG CCAAACCGCG GCTCCAGCCC TCTCCAGCCT CCGGCTCAGC     120

CGGCTCATCA GTCGGTCCGC GCCTTGCAGC TCCTCCAGAG GGACGCGCCC CGAG ATG       177
                                                            Met
                                                            -27

GAG AGC AAA GCC CTG CTC GTG CTG ACT CTG GCC GTG TGG CTC CAG AGT       225
Glu Ser Lys Ala Leu Leu Val Leu Thr Leu Ala Val Trp Leu Gln Ser
    -25                 -20                 -15

CTG ACC GCC TCC CGC GGA GGG GTG GCC GCC GCC GAC CAA AGA AGA GAT       273
Leu Thr Ala Ser Arg Gly Gly Val Ala Ala Ala Asp Gln Arg Arg Asp
-10                 -5                  1               5

TTT ATC GAC ATC GAA AGT AAA TTT GCC CTA AGG ACC CCT GAA GAC ACA       321
Phe Ile Asp Ile Glu Ser Lys Phe Ala Leu Arg Thr Pro Glu Asp Thr
            10                  15                  20

GCT GAG GAC ACT TGC CAC CTC ATT CCC GGA GTA GCA GAG TCC GTG GCT       369
Ala Glu Asp Thr Cys His Leu Ile Pro Gly Val Ala Glu Ser Val Ala
        25                  30                  35

ACC TGT CAT TTC AAT CAC AGC AGC AAA ACC TTC ATG GTG ATC CAT GGC       417
Thr Cys His Phe Asn His Ser Ser Lys Thr Phe Met Val Ile His Gly
    40                  45                  50

TGG ACG GTA ACA GGA ATG TAT GAG AGT TGG GTG CCA AAA CTT GTG GCC       465
Trp Thr Val Thr Gly Met Tyr Glu Ser Trp Val Pro Lys Leu Val Ala
55                  60                  65                  70
```

```
GCC CTG TAC AAG AGA GAA CCA GAC TCC AAT GTC ATT GTG GTG GAC TGG      513
Ala Leu Tyr Lys Arg Glu Pro Asp Ser Asn Val Ile Val Val Asp Trp
                75                  80                  85

CTG TCA CGG GCT CAG GAG CAT TAC CCA GTG TCC GCG GGC TAC ACC AAA      561
Leu Ser Arg Ala Gln Glu His Tyr Pro Val Ser Ala Gly Tyr Thr Lys
            90                  95                 100

CTG GTG GGA CAG GAT GTG GCC CGG TTT ATC AAC TGG ATG GAG GAG GAG      609
Leu Val Gly Gln Asp Val Ala Arg Phe Ile Asn Trp Met Glu Glu Glu
        105                 110                 115

TTT AAC TAC CCT CTG GAC AAT GTC CAT CTC TTG GGA TAC AGC CTT GGA      657
Phe Asn Tyr Pro Leu Asp Asn Val His Leu Leu Gly Tyr Ser Leu Gly
    120                 125                 130

GCC CAT GCT GCT GGC ATT GCA GGA AGT CTG ACC AAT AAG AAA GTC AAC      705
Ala His Ala Ala Gly Ile Ala Gly Ser Leu Thr Asn Lys Lys Val Asn
135                 140                 145                 150

AGA ATT ACT GGC CTC GAT CCA GCT GGA CCT AAC TTT GAG TAT GCA GAA      753
Arg Ile Thr Gly Leu Asp Pro Ala Gly Pro Asn Phe Glu Tyr Ala Glu
                155                 160                 165

GCC CCG AGT CGT CTT TCT CCT GAT GAT GCA GAT TTT GTA GAC GTC TTA      801
Ala Pro Ser Arg Leu Ser Pro Asp Asp Ala Asp Phe Val Asp Val Leu
            170                 175                 180

CAC ACA TTC ACC AGA GGG TCC CCT GGT CGA AGC ATT GGA ATC CAG AAA      849
His Thr Phe Thr Arg Gly Ser Pro Gly Arg Ser Ile Gly Ile Gln Lys
        185                 190                 195

CCA GTT GGG CAT GTT GAC ATT TAC CCG AAT GGA GGT ACT TTT CAG CCA      897
Pro Val Gly His Val Asp Ile Tyr Pro Asn Gly Gly Thr Phe Gln Pro
    200                 205                 210

GGA TGT AAC ATT GGA GAA GCT ATC CGC GTG ATT GCA GAG AGA GGA CTT      945
Gly Cys Asn Ile Gly Glu Ala Ile Arg Val Ile Ala Glu Arg Gly Leu
215                 220                 225                 230

GGA GAT GTG GAC CAG CTA GTG AAG TGC TCC CAC GAG CGC TCC ATT CAT      993
Gly Asp Val Asp Gln Leu Val Lys Cys Ser His Glu Arg Ser Ile His
                235                 240                 245

CTC TTC ATC GAC TCT CTG TTG AAT GAA GAA AAT CCA AGT AAG GCC TAC     1041
Leu Phe Ile Asp Ser Leu Leu Asn Glu Glu Asn Pro Ser Lys Ala Tyr
            250                 255                 260

AGG TGC AGT TCC AAG GAA GCC TTT GAG AAA GGG CTC TGC TTG AGT TGT     1089
Arg Cys Ser Ser Lys Glu Ala Phe Glu Lys Gly Leu Cys Leu Ser Cys
        265                 270                 275
```

```
AGA AAG AAC CGC TGC AAC AAT CTG GGC TAT GAG ATC AAT AAA GTC AGA      1137
Arg Lys Asn Arg Cys Asn Asn Leu Gly Tyr Glu Ile Asn Lys Val Arg
    280                 285                 290

GCC AAA AGA AGC AGC AAA ATG TAC CTG AAG ACT CGT TCT CAG ATG CCC      1185
Ala Lys Arg Ser Ser Lys Met Tyr Leu Lys Thr Arg Ser Gln Met Pro
295                 300                 305                 310

TAC AAA GTC TTC CAT TAC CAA GTA AAG ATT CAT TTT TCT GGG ACT GAG      1233
Tyr Lys Val Phe His Tyr Gln Val Lys Ile His Phe Ser Gly Thr Glu
                315                 320                 325

AGT GAA ACC CAT ACC AAT CAG GCC TTT GAG ATT TCT CTG TAT GGC ACC      1281
Ser Glu Thr His Thr Asn Gln Ala Phe Glu Ile Ser Leu Tyr Gly Thr
            330                 335                 340

GTG GCC GAG AGT GAG AAC ATC CCA TTC ACT CTG CCT GAA GTT TCC ACA      1329
Val Ala Glu Ser Glu Asn Ile Pro Phe Thr Leu Pro Glu Val Ser Thr
        345                 350                 355

AAT AAG ACC TAC TCC TTC CTA ATT TAC ACA GAG GTA GAT ATT GGA GAA      1377
Asn Lys Thr Tyr Ser Phe Leu Ile Tyr Thr Glu Val Asp Ile Gly Glu
    360                 365                 370

CTA CTC ATG TTG AAG CTC AAA TGG AAG AGT GAT TCA TAC TTT AGC TGG      1425
Leu Leu Met Leu Lys Leu Lys Trp Lys Ser Asp Ser Tyr Phe Ser Trp
375                 380                 385                 390

TCA GAC TGG TGG AGC AGT CCC GGC TTC GCC ATT CAG AAG ATC AGA GTA      1473
Ser Asp Trp Trp Ser Ser Pro Gly Phe Ala Ile Gln Lys Ile Arg Val
                395                 400                 405

AAA GCA GGA GAG ACT CAG AAA AAG GTG ATC TTC TGT TCT AGG GAG AAA      1521
Lys Ala Gly Glu Thr Gln Lys Lys Val Ile Phe Cys Ser Arg Glu Lys
            410                 415                 420

GTG TCT CAT TTG CAG AAA GGA AAG GCA CCT GCG GTA TTT GTG AAA TGC      1569
Val Ser His Leu Gln Lys Gly Lys Ala Pro Ala Val Phe Val Lys Cys
        425                 430                 435

CAT GAC AAG TCT CTG AAT AAG AAG TCA GGC TGA AACTGGGCGA ATCTACAGAA    1622
His Asp Lys Ser Leu Asn Lys Lys Ser Gly  *
    440                 445

CAAAGAACGG CATGTGAATT CTGTGAAGAA TGAAGTGGAG GAAGTAACTT TTACAAAACA    1682

TACCCAGTGT TTGGGGTGTT TCAAAAGTGG ATTTTCCTGA ATATTAATCC CAGCCCTACC    1742

CTTGTTAGTT ATTTTAGGAG ACAGTCTCAA GCACTAAAAA GTGGCTAATT CAATTTATGG    1802

GGTATAGTGG CCAAATAGCA CATCCTCCAA CGTTAAAAGA CAGTGGATCA TGAAAAGTGC    1862

TGTTTTGTCC TTTGAGAAAG AAATAATTGT TTGAGCGCAG AGTAAAATAA GGCTCCTTCA    1922
```

TGTGGCGTAT TGGGCCATAG CCTATAATTG GTTAGAACCT CCTATTTTAA TTGGAATTCT 1982

GGATCTTTCG GACTGAGGCC TTCTCAAACT TTACTCTAAG TCTCCAAGAA TACAGAAAAT 2042

GCTTTTCCGC GGCACGAATC AGACTCATCT ACACAGCAGT ATGAATGATG TTTTAGAATG 2102

ATTCCCTCTT GCTATTGGAA TGTGGTCCAG ACGTCAACCA GGAACATGTA ACTTGGAGAG 2162

GGACGAAGAA AGGGTCTGAT AAACACAGAG GTTTTAAACA GTCCCTACCA TTGGCCTGCA 2222

TCATGACAAA GTTACAAATT CAAGGAGATA TAAAATCTAG ATCAATTAAT TCTTAATAGG 2282

CTTTATCGTT TATTGCTTAA TCCCTCTCTC CCCCTTCTTT TTTGTCTCAA GATTATATTA 2342

TAATAATGTT CTCTGGGTAG GTGTTGAAAA TGAGCCTGTA ATCCTCAGCT GACACATAAT 2402

TTGAATGGTG CAGAAAAAAA AAAGATACCG TAATTTTATT ATTAGATTCT CCAAATGATT 2462

TTCATCAATT TAAAATCATT CAATATCTGA CAGTTACTCT TCAGTTTTAG GCTTACCTTG 2522

GTCCATGCTC AGTTGTACTT CCAGTGCGTC TCTTTTGTTC CTGGCTTTGA CATGAAAAGA 2582

TAGGTTTGAG TTCAAATTTT GCATTGTGTG AGCTTCTACA GATTTTAGAC AAGGACCGTT 2642

TTTACTAAGT AAAAGGGTGG AGAGGTTCCT GGGGTGGATT CCTAAGCAGT GCTTGTAAAC 2702

CATCGCGTGC AATGAGCCAG ATGGAGTACC ATGAGGGTTG TTATTTGTTG TTTTTAACAA 2762

CTAATCAAGA GTGAGTGAAC AACTATTTAT AAACTAGATC TCCTATTTTT CAGAATGCTC 2822

TTCTACGTAT AAATATGAAA TGATAAAGAT GTCAAATATC TCAGAGGCTA TAGCTGGGAA 2882

CCCGACTGTG AAAGTATGTG ATATCTGAAC ACATACTAGA AAGCTCTGCA TGTGTGTTGT 2942

CCTTCAGCAT AATTCGGAAG GGAAAACAGT CGATCAAGGG ATGTATTGGA ACATGTCGGA 3002

GTAGAAATTG TTCCTGATGT GCCAGAACTT CGACCCTTTC TCTGAGAGAG ATGATCGTGC 3062

CTATAAATAG TAGGACCAAT GTTGTGATTA ACATCATCAG GCTTGGAATG AATTCTCTCT 3122

AAAAATAAAA TGATGTATGA TTTGTTGTTG GCATCCCCTT TATTAATTCA TTAAATTTCT 3182

GGATTTGGGT TGTGACCCAG GGTGCATTAA CTTAAAAGAT TCACTAAAGC AGCACATAGC 3242

ACTGGGAACT CTGGCTCCGA AAAACTTTGT TATATATATC AAGGATGTTC TGGCTTTACA 3302

TTTTATTTAT TAGCTGTAAA TACATGTGTG GATGTGTAAA TGGAGCTTGT ACATATTGGA 3362

AAGGTCATTG TGGCTATCTG CATTTATAAA TGTGTGGTGC TAACTGTATG TGTCTTTATC 3422

AGTGATGGTC TCACAGAGCC AACTCACTCT TATGAAATGG GCTTTAACAA AACAAGAAAG 3482

AAACGTACTT AACTGTGTGA AGAAATGGAA TCAGCTTTTA ATAAAATTGA CAACATTTTA 3542

TTACCAC 3549

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 476 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Met Glu Ser Lys Ala Leu Leu Val Leu Thr Leu Ala Val Trp Leu Gln
-27      -25                 -20                 -15

Ser Leu Thr Ala Ser Arg Gly Gly Val Ala Ala Ala Asp Gln Arg Arg
    -10                 -5                  1               5

Asp Phe Ile Asp Ile Glu Ser Lys Phe Ala Leu Arg Thr Pro Glu Asp
                10                  15                  20

Thr Ala Glu Asp Thr Cys His Leu Ile Pro Gly Val Ala Glu Ser Val
            25                  30                  35

Ala Thr Cys His Phe Asn His Ser Ser Lys Thr Phe Met Val Ile His
        40                  45                  50

Gly Trp Thr Val Thr Gly Met Tyr Glu Ser Trp Val Pro Lys Leu Val
    55                  60                  65

Ala Ala Leu Tyr Lys Arg Glu Pro Asp Ser Asn Val Ile Val Val Asp
70                  75                  80                  85

Trp Leu Ser Arg Ala Gln Glu His Tyr Pro Val Ser Ala Gly Tyr Thr
                90                  95                  100

Lys Leu Val Gly Gln Asp Val Ala Arg Phe Ile Asn Trp Met Glu Glu
            105                 110                 115

Glu Phe Asn Tyr Pro Leu Asp Asn Val His Leu Leu Gly Tyr Ser Leu
        120                 125                 130

Gly Ala His Ala Ala Gly Ile Ala Gly Ser Leu Thr Asn Lys Lys Val
    135                 140                 145

Asn Arg Ile Thr Gly Leu Asp Pro Ala Gly Pro Asn Phe Glu Tyr Ala
150                 155                 160                 165

Glu Ala Pro Ser Arg Leu Ser Pro Asp Asp Ala Asp Phe Val Asp Val
                170                 175                 180

Leu His Thr Phe Thr Arg Gly Ser Pro Gly Arg Ser Ile Gly Ile Gln
            185                 190                 195
```

```
Lys Pro Val Gly His Val Asp Ile Tyr Pro Asn Gly Gly Thr Phe Gln
        200                 205                 210

Pro Gly Cys Asn Ile Gly Glu Ala Ile Arg Val Ile Ala Glu Arg Gly
    215                 220                 225

Leu Gly Asp Val Asp Gln Leu Val Lys Cys Ser His Glu Arg Ser Ile
230                 235                 240                 245

His Leu Phe Ile Asp Ser Leu Leu Asn Glu Glu Asn Pro Ser Lys Ala
                250                 255                 260

Tyr Arg Cys Ser Ser Lys Glu Ala Phe Glu Lys Gly Leu Cys Leu Ser
            265                 270                 275

Cys Arg Lys Asn Arg Cys Asn Asn Leu Gly Tyr Glu Ile Asn Lys Val
        280                 285                 290

Arg Ala Lys Arg Ser Ser Lys Met Tyr Leu Lys Thr Arg Ser Gln Met
    295                 300                 305

Pro Tyr Lys Val Phe His Tyr Gln Val Lys Ile His Phe Ser Gly Thr
310                 315                 320                 325

Glu Ser Glu Thr His Thr Asn Gln Ala Phe Glu Ile Ser Leu Tyr Gly
                330                 335                 340

Thr Val Ala Glu Ser Glu Asn Ile Pro Phe Thr Leu Pro Glu Val Ser
            345                 350                 355

Thr Asn Lys Thr Tyr Ser Phe Leu Ile Tyr Thr Glu Val Asp Ile Gly
        360                 365                 370

Glu Leu Leu Met Leu Lys Leu Lys Trp Lys Ser Asp Ser Tyr Phe Ser
    375                 380                 385

Trp Ser Asp Trp Trp Ser Ser Pro Gly Phe Ala Ile Gln Lys Ile Arg
390                 395                 400                 405

Val Lys Ala Gly Glu Thr Gln Lys Lys Val Ile Phe Cys Ser Arg Glu
                410                 415                 420

Lys Val Ser His Leu Gln Lys Gly Lys Ala Pro Ala Val Phe Val Lys
            425                 430                 435

Cys His Asp Lys Ser Leu Asn Lys Lys Ser Gly  *
        440                 445
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4020 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 1..109

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 23..3943

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION: 110..3940

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
GCCGAGCACC GCGCACCGCG TC ATG GGG GCC GCC TCG GGC CGC CGG GGG CCG        52
                         Met Gly Ala Ala Ser Gly Arg Arg Gly Pro
                         -29             -25                 -20

GGG CTG CTG CTG CCG CTG CCG CTG CTG TTG CTG CTG CCG CCG CAG CCC       100
Gly Leu Leu Leu Pro Leu Pro Leu Leu Leu Leu Leu Pro Pro Gln Pro
                -15                 -10                 -5

GCC CTG GCG TTG GAC CCC GGG CTG CAG CCC GGC AAC TTT TCT GCT GAC       148
Ala Leu Ala Leu Asp Pro Gly Leu Gln Pro Gly Asn Phe Ser Ala Asp
             1               5                  10

GAG GCC GGG GCG CAG CTC TTC GCG CAG AGC TAC AAC TCC AGC GCC GAA       196
Glu Ala Gly Ala Gln Leu Phe Ala Gln Ser Tyr Asn Ser Ser Ala Glu
     15                  20                  25

CAG GTG CTG TTC CAG AGC GTG GCC GCC AGC TGG GCG CAC GAC ACC AAC       244
Gln Val Leu Phe Gln Ser Val Ala Ala Ser Trp Ala His Asp Thr Asn
 30                  35                  40                  45

ATC ACC GCG GAG AAT GCA AGG CGC CAG GAG GAA GCA GCC CTG CTC AGC       292
Ile Thr Ala Glu Asn Ala Arg Arg Gln Glu Glu Ala Ala Leu Leu Ser
                 50                  55                  60

CAG GAG TTT GCG GAG GCC TGG GGC CAG AAG GCC AAG GAG CTG TAT GAA       340
Gln Glu Phe Ala Glu Ala Trp Gly Gln Lys Ala Lys Glu Leu Tyr Glu
             65                  70                  75

CCG ATC TGG CAG AAC TTC ACG GAC CCG CAG CTG CGC AGG ATC ATC GGA       388
Pro Ile Trp Gln Asn Phe Thr Asp Pro Gln Leu Arg Arg Ile Ile Gly
         80                  85                  90
```

```
GCT GTG CGA ACC CTG GGC TCT GCC AAC CTG CCC CTG GCT AAG CGG CAG        436
Ala Val Arg Thr Leu Gly Ser Ala Asn Leu Pro Leu Ala Lys Arg Gln
     95                  100                 105

CAG TAC AAC GCC CTG CTA AGC AAC ATG AGC AGG ATC TAC TCC ACC GCC        484
Gln Tyr Asn Ala Leu Leu Ser Asn Met Ser Arg Ile Tyr Ser Thr Ala
110                 115                 120                 125

AAG GTC TGC CTC CCC AAC AAG ACT GCC ACC TGC TGG TCC CTG GAC CCA        532
Lys Val Cys Leu Pro Asn Lys Thr Ala Thr Cys Trp Ser Leu Asp Pro
                130                 135                 140

GAT CTC ACC AAC ATC CTG GCT TCC TCG CGA AGC TAC GCC ATG CTC CTG        580
Asp Leu Thr Asn Ile Leu Ala Ser Ser Arg Ser Tyr Ala Met Leu Leu
            145                 150                 155

TTT GCC TGG GAG GGC TGG CAC AAC GCT GCG GGC ATC CCG CTG AAA CCG        628
Phe Ala Trp Glu Gly Trp His Asn Ala Ala Gly Ile Pro Leu Lys Pro
        160                 165                 170

CTG TAC GAG GAT TTC ACT GCC CTC AGC AAT GAA GCC TAC AAG CAG GAC        676
Leu Tyr Glu Asp Phe Thr Ala Leu Ser Asn Glu Ala Tyr Lys Gln Asp
    175                 180                 185

GGC TTC ACA GAC ACG GGG GCC TAC TGG CGC TCC TGG TAC AAC TCC CCC        724
Gly Phe Thr Asp Thr Gly Ala Tyr Trp Arg Ser Trp Tyr Asn Ser Pro
190                 195                 200                 205

ACC TTC GAG GAC GAT CTG GAA CAC CTC TAC CAA CAG CTA GAG CCC CTC        772
Thr Phe Glu Asp Asp Leu Glu His Leu Tyr Gln Gln Leu Glu Pro Leu
                210                 215                 220

TAC CTG AAC CTC CAT GCC TTC GTC CGC CGC GCA CTG CAT CGC CGA TAC        820
Tyr Leu Asn Leu His Ala Phe Val Arg Arg Ala Leu His Arg Arg Tyr
            225                 230                 235

GGA GAC AGA TAC ATC AAC CTC AGG GGA CCC ATC CCT GCT CAT CTG CTG        868
Gly Asp Arg Tyr Ile Asn Leu Arg Gly Pro Ile Pro Ala His Leu Leu
        240                 245                 250

GGA GAC ATG TGG GCC CAG AGC TGG GAA AAC ATC TAC GAC ATG GTG GTG        916
Gly Asp Met Trp Ala Gln Ser Trp Glu Asn Ile Tyr Asp Met Val Val
    255                 260                 265

CCT TTC CCA GAC AAG CCC AAC CTC GAT GTC ACC AGT ACT ATG CTG CAG        964
Pro Phe Pro Asp Lys Pro Asn Leu Asp Val Thr Ser Thr Met Leu Gln
270                 275                 280                 285

CAG GGC TGG AAC GCC ACG CAC ATG TTC CGG GTG GCA GAG GAG TTC TTC       1012
Gln Gly Trp Asn Ala Thr His Met Phe Arg Val Ala Glu Glu Phe Phe
                290                 295                 300
```

```
ACC TCC CTG GAG CTC TCC CCC ATG CCT CCC GAG TTC TGG GAA GGG TCG      1060
Thr Ser Leu Glu Leu Ser Pro Met Pro Pro Glu Phe Trp Glu Gly Ser
            305                 310                 315

ATG CTG GAG AAG CCG GCC GAC GGG CGG GAA GTG GTG TGC CAC GCC TCG      1108
Met Leu Glu Lys Pro Ala Asp Gly Arg Glu Val Val Cys His Ala Ser
        320                 325                 330

GCT TGG GAC TTC TAC AAC AGG AAA GAC TTC AGG ATC AAG CAG TGC ACA      1156
Ala Trp Asp Phe Tyr Asn Arg Lys Asp Phe Arg Ile Lys Gln Cys Thr
    335                 340                 345

CGG GTC ACG ATG GAC CAG CTC TCC ACA GTG CAC CAT GAG ATG GGC CAT      1204
Arg Val Thr Met Asp Gln Leu Ser Thr Val His His Glu Met Gly His
350                 355                 360                 365

ATA CAG TAC TAC CTG CAG TAC AAG GAT CTG CCC GTC TCC CTG CGT CGG      1252
Ile Gln Tyr Tyr Leu Gln Tyr Lys Asp Leu Pro Val Ser Leu Arg Arg
                370                 375                 380

GGG GCC AAC CCC GGC TTC CAT GAG GCC ATT GGG GAC GTG CTG GCG CTC      1300
Gly Ala Asn Pro Gly Phe His Glu Ala Ile Gly Asp Val Leu Ala Leu
            385                 390                 395

TCG GTC TCC ACT CCT GAA CAT CTG CAC AAA ATC GGC CTG CTG GAC CGT      1348
Ser Val Ser Thr Pro Glu His Leu His Lys Ile Gly Leu Leu Asp Arg
        400                 405                 410

GTC ACC AAT GAC ACG GAA AGT GAC ATC AAT TAC TTG CTA AAA ATG GCA      1396
Val Thr Asn Asp Thr Glu Ser Asp Ile Asn Tyr Leu Leu Lys Met Ala
    415                 420                 425

CTG GAA AAA ATT GCC TTC CTG CCC TTT GGC TAC TTG GTG GAC CAG TGG      1444
Leu Glu Lys Ile Ala Phe Leu Pro Phe Gly Tyr Leu Val Asp Gln Trp
430                 435                 440                 445

CGC TGG GGG GTC TTT AGT GGG CGT ACC CCC CCT TCC CGC TAC AAC TTC      1492
Arg Trp Gly Val Phe Ser Gly Arg Thr Pro Pro Ser Arg Tyr Asn Phe
                450                 455                 460

GAC TGG TGG TAT CTT CGA ACC AAG TAT CAG GGG ATC TGT CCT CCT GTT      1540
Asp Trp Trp Tyr Leu Arg Thr Lys Tyr Gln Gly Ile Cys Pro Pro Val
            465                 470                 475

ACC CGA AAC GAA ACC CAC TTT GAT GCT GGA GCT AAG TTT CAT GTT CCA      1588
Thr Arg Asn Glu Thr His Phe Asp Ala Gly Ala Lys Phe His Val Pro
        480                 485                 490

AAT GTG ACA CCA TAC ATC AGG TAC TTT GTG AGT TTT GTC CTG CAG TTC      1636
Asn Val Thr Pro Tyr Ile Arg Tyr Phe Val Ser Phe Val Leu Gln Phe
    495                 500                 505
```

```
CAG TTC CAT GAA GCC CTG TGC AAG GAG GCA GGC TAT GAG GGC CCA CTG      1684
Gln Phe His Glu Ala Leu Cys Lys Glu Ala Gly Tyr Glu Gly Pro Leu
510                 515                 520                 525

CAC CAG TGT GAC ATC TAC CGG TCC ACC AAG GCA GGG GCC AAG CTC CGG      1732
His Gln Cys Asp Ile Tyr Arg Ser Thr Lys Ala Gly Ala Lys Leu Arg
                530                 535                 540

AAG GTG CTG CAG GCT GGC TCC TCC AGG CCC TGG CAG GAG GTG CTG AAG      1780
Lys Val Leu Gln Ala Gly Ser Ser Arg Pro Trp Gln Glu Val Leu Lys
            545                 550                 555

GAC ATG GTC GGC TTA GAT GCC CTG GAT GCC CAG CCG CTG CTC AAG TAC      1828
Asp Met Val Gly Leu Asp Ala Leu Asp Ala Gln Pro Leu Leu Lys Tyr
        560                 565                 570

TTC CAG CCA GTC ACC CAG TGG CTG CAG GAG CAG AAC CAG CAG AAC GGC      1876
Phe Gln Pro Val Thr Gln Trp Leu Gln Glu Gln Asn Gln Gln Asn Gly
    575                 580                 585

GAG GTC CTG GGC TGG CCC GAG TAC CAG TGG CAC CCG CCG TTG CCT GAC      1924
Glu Val Leu Gly Trp Pro Glu Tyr Gln Trp His Pro Pro Leu Pro Asp
590                 595                 600                 605

AAC TAC CCG GAG GGC ATA GAC CTG GTG ACT GAT GAG GCT GAG GCC AGC      1972
Asn Tyr Pro Glu Gly Ile Asp Leu Val Thr Asp Glu Ala Glu Ala Ser
                610                 615                 620

AAG TTT GTG GAG GAA TAT GAC CGG ACA TCC CAG GTG GTG TGG AAC GAG      2020
Lys Phe Val Glu Glu Tyr Asp Arg Thr Ser Gln Val Val Trp Asn Glu
            625                 630                 635

TAT GCC GAG GCC AAC TGG AAC TAC AAC ACC AAC ATC ACC ACA GAG ACC      2068
Tyr Ala Glu Ala Asn Trp Asn Tyr Asn Thr Asn Ile Thr Thr Glu Thr
        640                 645                 650

AGC AAG ATT CTG CTG CAG AAG AAC ATG CAA ATA GCC AAC CAC ACC CTG      2116
Ser Lys Ile Leu Leu Gln Lys Asn Met Gln Ile Ala Asn His Thr Leu
    655                 660                 665

AAG TAC GGC ACC CAG GCC AGG AAG TTT GAT GTG AAC CAG TTG CAG AAC      2164
Lys Tyr Gly Thr Gln Ala Arg Lys Phe Asp Val Asn Gln Leu Gln Asn
670                 675                 680                 685

ACC ACT ATC AAG CGG ATC ATA AAG AAG GTT CAG GAC CTA GAA CGG GCA      2212
Thr Thr Ile Lys Arg Ile Ile Lys Lys Val Gln Asp Leu Glu Arg Ala
                690                 695                 700

GCG CTG CCT GCC CAG GAG CTG GAG GAG TAC AAC AAG ATC CTG TTG GAT      2260
Ala Leu Pro Ala Gln Glu Leu Glu Glu Tyr Asn Lys Ile Leu Leu Asp
            705                 710                 715
```

```
ATG GAA ACC ACC TAC AGC GTG GCC ACT GTG TGC CAC CCG AAT GGC AGC      2308
Met Glu Thr Thr Tyr Ser Val Ala Thr Val Cys His Pro Asn Gly Ser
        720                 725                 730

TGC CTG CAG CTC GAG CCA GAT CTG ACG AAT GTG ATG GCC ACA TCC CGG      2356
Cys Leu Gln Leu Glu Pro Asp Leu Thr Asn Val Met Ala Thr Ser Arg
    735                 740                 745

AAA TAT GAA GAC CTG TTA TGG GCA TGG GAG GGC TGG CGA GAC AAG GCG      2404
Lys Tyr Glu Asp Leu Leu Trp Ala Trp Glu Gly Trp Arg Asp Lys Ala
750                 755                 760                 765

GGG AGA GCC ATC CTC CAG TTT TAC CCG AAA TAC GTG GAA CTC ATC AAC      2452
Gly Arg Ala Ile Leu Gln Phe Tyr Pro Lys Tyr Val Glu Leu Ile Asn
                770                 775                 780

CAG GCT GCC CGG CTC AAT GGC TAT GTA GAT GCA GGG GAC TCG TGG AGG      2500
Gln Ala Ala Arg Leu Asn Gly Tyr Val Asp Ala Gly Asp Ser Trp Arg
            785                 790                 795

TCT ATG TAC GAG ACA CCA TCC CTG GAG CAA GAC CTG GAG CGG CTC TTC      2548
Ser Met Tyr Glu Thr Pro Ser Leu Glu Gln Asp Leu Glu Arg Leu Phe
        800                 805                 810

CAG GAG CTG CAG CCA CTC TAC CTC AAC CTG CAT GCC TAC GTG CGC CGG      2596
Gln Glu Leu Gln Pro Leu Tyr Leu Asn Leu His Ala Tyr Val Arg Arg
    815                 820                 825

GCC CTG CAC CGT CAC TAC GGG GCC CAG CAC ATC AAC CTG GAG GGG CCC      2644
Ala Leu His Arg His Tyr Gly Ala Gln His Ile Asn Leu Glu Gly Pro
830                 835                 840                 845

ATT CCT GCT CAC CTG CTG GGG AAC ATG TGG GCG CAG ACC TGG TCC AAC      2692
Ile Pro Ala His Leu Leu Gly Asn Met Trp Ala Gln Thr Trp Ser Asn
                850                 855                 860

ATC TAT GAC TTG GTG GTG CCC TTC CCT TCA GCC CCC TCG ATG GAC ACC      2740
Ile Tyr Asp Leu Val Val Pro Phe Pro Ser Ala Pro Ser Met Asp Thr
            865                 870                 875

ACA GAG GCT ATG CTA AAG CAG GGC TGG ACG CCC AGG AGG ATG TTT AAG      2788
Thr Glu Ala Met Leu Lys Gln Gly Trp Thr Pro Arg Arg Met Phe Lys
        880                 885                 890

GAG GCT GAT GAT TTC TTC ACC TCC CTG GGG CTG CTG CCC GTG CCT CCT      2836
Glu Ala Asp Asp Phe Phe Thr Ser Leu Gly Leu Leu Pro Val Pro Pro
    895                 900                 905

GAG TTC TGG AAC AAG TCG ATG CTG GAG AAG CCA ACC GAC GGG CGG GAG      2884
Glu Phe Trp Asn Lys Ser Met Leu Glu Lys Pro Thr Asp Gly Arg Glu
910                 915                 920                 925
```

```
GTG GTC TGC CAC GCC TCG GCC TGG GAC TTC TAC AAC GGC AAG GAC TTC      2932
Val Val Cys His Ala Ser Ala Trp Asp Phe Tyr Asn Gly Lys Asp Phe
                930                 935                 940

CGG ATC AAG CAG TGC ACC ACC GTG AAC TTG GAG GAC CTG GTG GTG GCC      2980
Arg Ile Lys Gln Cys Thr Thr Val Asn Leu Glu Asp Leu Val Val Ala
            945                 950                 955

CAC CAC GAA ATG GGC CAC ATC CAG TAT TTC ATG CAG TAC AAA GAC TTA      3028
His His Glu Met Gly His Ile Gln Tyr Phe Met Gln Tyr Lys Asp Leu
        960                 965                 970

CCT GTG GCC TTG AGG GAG GGT GCC AAC CCC GGC TTC CAT GAG GCC ATT      3076
Pro Val Ala Leu Arg Glu Gly Ala Asn Pro Gly Phe His Glu Ala Ile
    975                 980                 985

GGG GAC GTG CTA GCC CTC TCA GTG TCT ACG CCC AAG CAC CTG CAC AGT      3124
Gly Asp Val Leu Ala Leu Ser Val Ser Thr Pro Lys His Leu His Ser
990                 995                 1000                1005

CTC AAC CTG CTG AGC AGT GAG GGT GGC AGC GAC GAG CAT GAC ATC AAC      3172
Leu Asn Leu Leu Ser Ser Glu Gly Gly Ser Asp Glu His Asp Ile Asn
                1010                1015                1020

TTT CTG ATG AAG ATG GCC CTT GAC AAG ATC GCC TTT ATC CCC TTC AGC      3220
Phe Leu Met Lys Met Ala Leu Asp Lys Ile Ala Phe Ile Pro Phe Ser
            1025                1030                1035

TAC CTC GTC GAT CAG TGG CGC TGG AGG GTA TTT GAT GGA AGC ATC ACC      3268
Tyr Leu Val Asp Gln Trp Arg Trp Arg Val Phe Asp Gly Ser Ile Thr
        1040                1045                1050

AAG GAG AAC TAT AAC CAG GAG TGG TGG AGC CTC AGG CTG AAG TAC CAG      3316
Lys Glu Asn Tyr Asn Gln Glu Trp Trp Ser Leu Arg Leu Lys Tyr Gln
    1055                1060                1065

GGC CTC TGC CCC CCA GTG CCC AGG ACT CAA GGT GAC TTT GAC CCA GGG      3364
Gly Leu Cys Pro Pro Val Pro Arg Thr Gln Gly Asp Phe Asp Pro Gly
1070                1075                1080                1085

GCC AAG TTC CAC ATT CCT TCT AGC GTG CCT TAC ATC AGG TAC TTT GTC      3412
Ala Lys Phe His Ile Pro Ser Ser Val Pro Tyr Ile Arg Tyr Phe Val
                1090                1095                1100

AGC TTC ATC ATC CAG TTC CAG TTC CAC GAG GCA CTG TGC CAG GCA GCT      3460
Ser Phe Ile Ile Gln Phe Gln Phe His Glu Ala Leu Cys Gln Ala Ala
            1105                1110                1115

GGC CAC ACG GGC CCC CTG CAC AAG TGT GAC ATC TAC CAG TCC AAG GAG      3508
Gly His Thr Gly Pro Leu His Lys Cys Asp Ile Tyr Gln Ser Lys Glu
        1120                1125                1130
```

```
GCC GGG CAG CGC CTG GCG ACC GCC ATG AAG CTG GGC TTC AGT AGG CCG      3556
Ala Gly Gln Arg Leu Ala Thr Ala Met Lys Leu Gly Phe Ser Arg Pro
    1135                1140                1145

TGG CCG GAA GCC ATG CAG CTG ATC ACG GGC CAG CCC AAC ATG AGC GCC      3604
Trp Pro Glu Ala Met Gln Leu Ile Thr Gly Gln Pro Asn Met Ser Ala
1150                1155                1160                1165

TCG GCC ATG TTG AGC TAC TTC AAG CCG CTG CTG GAC TGG CTC CGC ACG      3652
Ser Ala Met Leu Ser Tyr Phe Lys Pro Leu Leu Asp Trp Leu Arg Thr
                1170                1175                1180

GAG AAC GAG CTG CAT GGG GAG AAG CTG GGC TGG CCG CAG TAC AAC TGG      3700
Glu Asn Glu Leu His Gly Glu Lys Leu Gly Trp Pro Gln Tyr Asn Trp
            1185                1190                1195

ACG CCG AAC TCC GCT CGC TCA GAA GGG CCC CTC CCA GAC AGC GGC CGC      3748
Thr Pro Asn Ser Ala Arg Ser Glu Gly Pro Leu Pro Asp Ser Gly Arg
        1200                1205                1210

GTC AGC TTC CTG GGC CTG GAC CTG GAT GCG CAG CAG GCC CGC GTG GGC      3796
Val Ser Phe Leu Gly Leu Asp Leu Asp Ala Gln Gln Ala Arg Val Gly
    1215                1220                1225

CAG TGG CTG CTG CTC TTC CTG GGC ATC GCC CTG CTG GTA GCC ACC CTG      3844
Gln Trp Leu Leu Leu Phe Leu Gly Ile Ala Leu Leu Val Ala Thr Leu
1230                1235                1240                1245

GGC CTC AGC CAG CGG CTC TTC AGC ATC CGC CAC CGC AGC CTC CAC CGG      3892
Gly Leu Ser Gln Arg Leu Phe Ser Ile Arg His Arg Ser Leu His Arg
                1250                1255                1260

CAC TCC CAC GGG CCC CAG TTC GGC TCC GAG GTG GAG CTG AGA CAC TCC      3940
His Ser His Gly Pro Gln Phe Gly Ser Glu Val Glu Leu Arg His Ser
            1265                1270                1275

TGA GGTGACCCGG CTGGGTCGGC CCTGCCCAAG GGCCTCCCAC CAGAGACTGG          3993
*

GATGGGAACA CTGGTGGGCA GCTGAGG                                       4020
```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1307 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
Met Gly Ala Ala Ser Gly Arg Arg Gly Pro Gly Leu Leu Leu Pro Leu
-29              -25                 -20                 -15

Pro Leu Leu Leu Leu Leu Pro Pro Gln Pro Ala Leu Ala Leu Asp Pro
             -10                 -5                  1

Gly Leu Gln Pro Gly Asn Phe Ser Ala Asp Glu Ala Gly Ala Gln Leu
      5                  10                  15

Phe Ala Gln Ser Tyr Asn Ser Ser Ala Glu Gln Val Leu Phe Gln Ser
 20                  25                  30                  35

Val Ala Ala Ser Trp Ala His Asp Thr Asn Ile Thr Ala Glu Asn Ala
                 40                  45                  50

Arg Arg Gln Glu Glu Ala Ala Leu Leu Ser Gln Glu Phe Ala Glu Ala
              55                  60                  65

Trp Gly Gln Lys Ala Lys Glu Leu Tyr Glu Pro Ile Trp Gln Asn Phe
          70                  75                  80

Thr Asp Pro Gln Leu Arg Arg Ile Ile Gly Ala Val Arg Thr Leu Gly
      85                  90                  95

Ser Ala Asn Leu Pro Leu Ala Lys Arg Gln Gln Tyr Asn Ala Leu Leu
100                 105                 110                 115

Ser Asn Met Ser Arg Ile Tyr Ser Thr Ala Lys Val Cys Leu Pro Asn
                120                 125                 130

Lys Thr Ala Thr Cys Trp Ser Leu Asp Pro Asp Leu Thr Asn Ile Leu
             135                 140                 145

Ala Ser Ser Arg Ser Tyr Ala Met Leu Leu Phe Ala Trp Glu Gly Trp
         150                 155                 160

His Asn Ala Ala Gly Ile Pro Leu Lys Pro Leu Tyr Glu Asp Phe Thr
     165                 170                 175

Ala Leu Ser Asn Glu Ala Tyr Lys Gln Asp Gly Phe Thr Asp Thr Gly
180                 185                 190                 195

Ala Tyr Trp Arg Ser Trp Tyr Asn Ser Pro Thr Phe Glu Asp Asp Leu
                200                 205                 210

Glu His Leu Tyr Gln Gln Leu Glu Pro Leu Tyr Leu Asn Leu His Ala
             215                 220                 225

Phe Val Arg Arg Ala Leu His Arg Arg Tyr Gly Asp Arg Tyr Ile Asn
         230                 235                 240

Leu Arg Gly Pro Ile Pro Ala His Leu Leu Gly Asp Met Trp Ala Gln
     245                 250                 255
```

```
Ser Trp Glu Asn Ile Tyr Asp Met Val Val Pro Phe Pro Asp Lys Pro
260                 265                 270                 275

Asn Leu Asp Val Thr Ser Thr Met Leu Gln Gln Gly Trp Asn Ala Thr
                280                 285                 290

His Met Phe Arg Val Ala Glu Glu Phe Phe Thr Ser Leu Glu Leu Ser
            295                 300                 305

Pro Met Pro Pro Glu Phe Trp Glu Gly Ser Met Leu Glu Lys Pro Ala
        310                 315                 320

Asp Gly Arg Glu Val Val Cys His Ala Ser Ala Trp Asp Phe Tyr Asn
    325                 330                 335

Arg Lys Asp Phe Arg Ile Lys Gln Cys Thr Arg Val Thr Met Asp Gln
340                 345                 350                 355

Leu Ser Thr Val His His Glu Met Gly His Ile Gln Tyr Tyr Leu Gln
                360                 365                 370

Tyr Lys Asp Leu Pro Val Ser Leu Arg Arg Gly Ala Asn Pro Gly Phe
            375                 380                 385

His Glu Ala Ile Gly Asp Val Leu Ala Leu Ser Val Ser Thr Pro Glu
        390                 395                 400

His Leu His Lys Ile Gly Leu Leu Asp Arg Val Thr Asn Asp Thr Glu
    405                 410                 415

Ser Asp Ile Asn Tyr Leu Leu Lys Met Ala Leu Glu Lys Ile Ala Phe
420                 425                 430                 435

Leu Pro Phe Gly Tyr Leu Val Asp Gln Trp Arg Trp Gly Val Phe Ser
                440                 445                 450

Gly Arg Thr Pro Pro Ser Arg Tyr Asn Phe Asp Trp Trp Tyr Leu Arg
            455                 460                 465

Thr Lys Tyr Gln Gly Ile Cys Pro Pro Val Thr Arg Asn Glu Thr His
        470                 475                 480

Phe Asp Ala Gly Ala Lys Phe His Val Pro Asn Val Thr Pro Tyr Ile
    485                 490                 495

Arg Tyr Phe Val Ser Phe Val Leu Gln Phe Gln Phe His Glu Ala Leu
500                 505                 510                 515

Cys Lys Glu Ala Gly Tyr Glu Gly Pro Leu His Gln Cys Asp Ile Tyr
                520                 525                 530

Arg Ser Thr Lys Ala Gly Ala Lys Leu Arg Lys Val Leu Gln Ala Gly
            535                 540                 545
```

```
Ser Ser Arg Pro Trp Gln Glu Val Leu Lys Asp Met Val Gly Leu Asp
        550                 555                 560

Ala Leu Asp Ala Gln Pro Leu Leu Lys Tyr Phe Gln Pro Val Thr Gln
    565                 570                 575

Trp Leu Gln Glu Gln Asn Gln Gln Asn Gly Glu Val Leu Gly Trp Pro
580                 585                 590                 595

Glu Tyr Gln Trp His Pro Pro Leu Pro Asp Asn Tyr Pro Glu Gly Ile
                600                 605                 610

Asp Leu Val Thr Asp Glu Ala Glu Ala Ser Lys Phe Val Glu Glu Tyr
            615                 620                 625

Asp Arg Thr Ser Gln Val Val Trp Asn Glu Tyr Ala Glu Ala Asn Trp
        630                 635                 640

Asn Tyr Asn Thr Asn Ile Thr Thr Glu Thr Ser Lys Ile Leu Leu Gln
    645                 650                 655

Lys Asn Met Gln Ile Ala Asn His Thr Leu Lys Tyr Gly Thr Gln Ala
660                 665                 670                 675

Arg Lys Phe Asp Val Asn Gln Leu Gln Asn Thr Thr Ile Lys Arg Ile
                680                 685                 690

Ile Lys Lys Val Gln Asp Leu Glu Arg Ala Ala Leu Pro Ala Gln Glu
            695                 700                 705

Leu Glu Glu Tyr Asn Lys Ile Leu Leu Asp Met Glu Thr Thr Tyr Ser
        710                 715                 720

Val Ala Thr Val Cys His Pro Asn Gly Ser Cys Leu Gln Leu Glu Pro
    725                 730                 735

Asp Leu Thr Asn Val Met Ala Thr Ser Arg Lys Tyr Glu Asp Leu Leu
740                 745                 750                 755

Trp Ala Trp Glu Gly Trp Arg Asp Lys Ala Gly Arg Ala Ile Leu Gln
                760                 765                 770

Phe Tyr Pro Lys Tyr Val Glu Leu Ile Asn Gln Ala Ala Arg Leu Asn
            775                 780                 785

Gly Tyr Val Asp Ala Gly Asp Ser Trp Arg Ser Met Tyr Glu Thr Pro
        790                 795                 800

Ser Leu Glu Gln Asp Leu Glu Arg Leu Phe Gln Glu Leu Gln Pro Leu
    805                 810                 815

Tyr Leu Asn Leu His Ala Tyr Val Arg Arg Ala Leu His Arg His Tyr
820                 825                 830                 835
```

Gly Ala Gln His Ile Asn Leu Glu Gly Pro Ile Pro Ala His Leu Leu
840 845 850

Gly Asn Met Trp Ala Gln Thr Trp Ser Asn Ile Tyr Asp Leu Val Val
855 860 865

Pro Phe Pro Ser Ala Pro Ser Met Asp Thr Thr Glu Ala Met Leu Lys
870 875 880

Gln Gly Trp Thr Pro Arg Arg Met Phe Lys Glu Ala Asp Asp Phe Phe
885 890 895

Thr Ser Leu Gly Leu Leu Pro Val Pro Pro Glu Phe Trp Asn Lys Ser
900 905 910 915

Met Leu Glu Lys Pro Thr Asp Gly Arg Glu Val Val Cys His Ala Ser
920 925 930

Ala Trp Asp Phe Tyr Asn Gly Lys Asp Phe Arg Ile Lys Gln Cys Thr
935 940 945

Thr Val Asn Leu Glu Asp Leu Val Val Ala His His Glu Met Gly His
950 955 960

Ile Gln Tyr Phe Met Gln Tyr Lys Asp Leu Pro Val Ala Leu Arg Glu
965 970 975

Gly Ala Asn Pro Gly Phe His Glu Ala Ile Gly Asp Val Leu Ala Leu
980 985 990 995

Ser Val Ser Thr Pro Lys His Leu His Ser Leu Asn Leu Leu Ser Ser
1000 1005 1010

Glu Gly Gly Ser Asp Glu His Asp Ile Asn Phe Leu Met Lys Met Ala
1015 1020 1025

Leu Asp Lys Ile Ala Phe Ile Pro Phe Ser Tyr Leu Val Asp Gln Trp
1030 1035 1040

Arg Trp Arg Val Phe Asp Gly Ser Ile Thr Lys Glu Asn Tyr Asn Gln
1045 1050 1055

Glu Trp Trp Ser Leu Arg Leu Lys Tyr Gln Gly Leu Cys Pro Pro Val
1060 1065 1070 1075

Pro Arg Thr Gln Gly Asp Phe Asp Pro Gly Ala Lys Phe His Ile Pro
1080 1085 1090

Ser Ser Val Pro Tyr Ile Arg Tyr Phe Val Ser Phe Ile Ile Gln Phe
1095 1100 1105

Gln Phe His Glu Ala Leu Cys Gln Ala Ala Gly His Thr Gly Pro Leu
1110 1115 1120

```
His Lys Cys Asp Ile Tyr Gln Ser Lys Glu Ala Gly Gln Arg Leu Ala
    1125                1130                1135

Thr Ala Met Lys Leu Gly Phe Ser Arg Pro Trp Pro Glu Ala Met Gln
1140                1145                1150                1155

Leu Ile Thr Gly Gln Pro Asn Met Ser Ala Ser Ala Met Leu Ser Tyr
                1160                1165                1170

Phe Lys Pro Leu Leu Asp Trp Leu Arg Thr Glu Asn Glu Leu His Gly
            1175                1180                1185

Glu Lys Leu Gly Trp Pro Gln Tyr Asn Trp Thr Pro Asn Ser Ala Arg
        1190                1195                1200

Ser Glu Gly Pro Leu Pro Asp Ser Gly Arg Val Ser Phe Leu Gly Leu
    1205                1210                1215

Asp Leu Asp Ala Gln Gln Ala Arg Val Gly Gln Trp Leu Leu Leu Phe
1220                1225                1230                1235

Leu Gly Ile Ala Leu Leu Val Ala Thr Leu Gly Leu Ser Gln Arg Leu
                1240                1245                1250

Phe Ser Ile Arg His Arg Ser Leu His Arg His Ser His Gly Pro Gln
            1255                1260                1265

Phe Gly Ser Glu Val Glu Leu Arg His Ser  *
        1270                1275
```

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CCGTTTGTGC AGGGCCTGGC TCTCT 25

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CAGGGTGCTG TCCACACTGG ACCCC 25

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TTTAGGCCTG AAGTTTCCAC 20

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CTCCCTAGAA GAGAAGATC 19

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TAGGAGGTTG AGGCACCTGT GC 22

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GTGGGTGAAT CACCTGAGGT C 21

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4604 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: mRNA
(B) LOCATION: -1..4604

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 116..1399

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

```
GGAACAGCTT GTCCACCCGC CGGCCGGACC AGAAGCCTTT GGGTCTGAAG TGTCTGTGAG      60

ACCTCACAGA AGAGCACCCC TGGGCTCCAC TTACCTGCCC CCTGCTCCTT CAGGG ATG      118
                                                             Met

GAG GCA ATG GCG GCC AGC ACT TCC CTG CCT GAC CCT GGA GAC TTT GAC       166
Glu Ala Met Ala Ala Ser Thr Ser Leu Pro Asp Pro Gly Asp Phe Asp
1280                 1285                 1290                 1295

CGG AAC GTG CCC CGG ATC TGT GGG GTG TGT GGA GAC CGA GCC ACT GGC       214
Arg Asn Val Pro Arg Ile Cys Gly Val Cys Gly Asp Arg Ala Thr Gly
                1300                 1305                 1310

TTT CAC TTC AAT GCT ATG ACC TGT GAA GGC TGC AAA GGC TTC TTC AGG       262
Phe His Phe Asn Ala Met Thr Cys Glu Gly Cys Lys Gly Phe Phe Arg
            1315                 1320                 1325

CGA AGC ATG AAG CGG AAG GCA CTA TTC ACC TGC CCC TTC AAC GGG GAC       310
Arg Ser Met Lys Arg Lys Ala Leu Phe Thr Cys Pro Phe Asn Gly Asp
        1330                 1335                 1340

TGC CGC ATC ACC AAG GAC AAC CGA CGC CAC TGC CAG GCC TGC CGG CTC       358
Cys Arg Ile Thr Lys Asp Asn Arg Arg His Cys Gln Ala Cys Arg Leu
    1345                 1350                 1355

AAA CGC TGT GTG GAC ATC GGC ATG ATG AAG GAG TTC ATT CTG ACA GAT       406
Lys Arg Cys Val Asp Ile Gly Met Met Lys Glu Phe Ile Leu Thr Asp
1360                 1365                 1370                 1375

GAG GAA GTG CAG AGG AAG CGG GAG ATG ATC CTG AAG CGG AAG GAG GAG       454
Glu Glu Val Gln Arg Lys Arg Glu Met Ile Leu Lys Arg Lys Glu Glu
                1380                 1385                 1390

GAG GCC TTG AAG GAC AGT CTG CGG CCC AAG CTG TCT GAG GAG CAG CAG       502
Glu Ala Leu Lys Asp Ser Leu Arg Pro Lys Leu Ser Glu Glu Gln Gln
            1395                 1400                 1405
```

```
CGC ATC ATT GCC ATA CTG CTG GAC GCC CAC CAT AAG ACC TAC GAC CCC      550
Arg Ile Ile Ala Ile Leu Leu Asp Ala His His Lys Thr Tyr Asp Pro
        1410                1415                1420

ACC TAC TCC GAC TTC TGC CAG TTC CGG CCT CCA GTT CGT GTG AAT GAT      598
Thr Tyr Ser Asp Phe Cys Gln Phe Arg Pro Pro Val Arg Val Asn Asp
    1425                1430                1435

GGT GGA GGG AGC CAT CCT TCC AGG CCC AAC TCC AGA CAC ACT CCC AGC      646
Gly Gly Gly Ser His Pro Ser Arg Pro Asn Ser Arg His Thr Pro Ser
1440                1445                1450                1455

TTC TCT GGG GAC TCC TCC TCC TCC TGC TCA GAT CAC TGT ATC ACC TCT      694
Phe Ser Gly Asp Ser Ser Ser Ser Cys Ser Asp His Cys Ile Thr Ser
                1460                1465                1470

TCA GAC ATG ATG GAC TCG TCC AGC TTC TCC AAT CTG GAT CTG AGT GAA      742
Ser Asp Met Met Asp Ser Ser Ser Phe Ser Asn Leu Asp Leu Ser Glu
            1475                1480                1485

GAA GAT TCA GAT GAC CCT TCT GTG ACC CTA GAG CTG TCC CAG CTC TCC      790
Glu Asp Ser Asp Asp Pro Ser Val Thr Leu Glu Leu Ser Gln Leu Ser
        1490                1495                1500

ATG CTG CCC CAC CTG GCT GAC CTG GTC AGT TAC AGC ATC CAA AAG GTC      838
Met Leu Pro His Leu Ala Asp Leu Val Ser Tyr Ser Ile Gln Lys Val
    1505                1510                1515

ATT GGC TTT GCT AAG ATG ATA CCA GGA TTC AGA GAC CTC ACC TCT GAG      886
Ile Gly Phe Ala Lys Met Ile Pro Gly Phe Arg Asp Leu Thr Ser Glu
1520                1525                1530                1535

GAC CAG ATC GTA CTG CTG AAG TCA AGT GCC ATT GAG GTC ATC ATG TTG      934
Asp Gln Ile Val Leu Leu Lys Ser Ser Ala Ile Glu Val Ile Met Leu
                1540                1545                1550

CGC TCC AAT GAG TCC TTC ACC ATG GAC GAC ATG TCC TGG ACC TGT GGC      982
Arg Ser Asn Glu Ser Phe Thr Met Asp Asp Met Ser Trp Thr Cys Gly
            1555                1560                1565

AAC CAA GAC TAC AAG TAC CGC GTC AGT GAC GTG ACC AAA GCC GGA CAC     1030
Asn Gln Asp Tyr Lys Tyr Arg Val Ser Asp Val Thr Lys Ala Gly His
        1570                1575                1580

AGC CTG GAG CTG ATT GAG CCC CTC ATC AAG TTC CAG GTG GGA CTG AAG     1078
Ser Leu Glu Leu Ile Glu Pro Leu Ile Lys Phe Gln Val Gly Leu Lys
    1585                1590                1595

AAG CTG AAC TTG CAT GAG GAG GAG CAT GTC CTG CTC ATG GCC ATC TGC     1126
Lys Leu Asn Leu His Glu Glu Glu His Val Leu Leu Met Ala Ile Cys
1600                1605                1610                1615

ATC GTC TCC CCA GAT CGT CCT GGG GTG CAG GAC GCC GCG CTG ATT GAG     1174
Ile Val Ser Pro Asp Arg Pro Gly Val Gln Asp Ala Ala Leu Ile Glu
                1620                1625                1630
```

```
GCC ATC CAG GAC CGC CTG TCC AAC ACA CTG CAG ACG TAC ATC CGC TGC      1222
Ala Ile Gln Asp Arg Leu Ser Asn Thr Leu Gln Thr Tyr Ile Arg Cys
            1635                1640                1645

CGC CAC CCG CCC CCG GGC AGC CAC CTG CTC TAT GCC AAG ATG ATC CAG      1270
Arg His Pro Pro Pro Gly Ser His Leu Leu Tyr Ala Lys Met Ile Gln
        1650                1655                1660

AAG CTA GCC GAC CTG CGC AGC CTC AAT GAG GAG CAC TCC AAG CAG TAC      1318
Lys Leu Ala Asp Leu Arg Ser Leu Asn Glu Glu His Ser Lys Gln Tyr
    1665                1670                1675

CGC TGC CTC TCC TTC CAG CCT GAG TGC AGC ATG AAG CTA ACG CCC CTT      1366
Arg Cys Leu Ser Phe Gln Pro Glu Cys Ser Met Lys Leu Thr Pro Leu
1680                1685                1690                1695

GTG CTC GAA GTG TTT GGC AAT GAG ATC TCC TGA CTAGGACAGC CTGTGCGGTG    1419
Val Leu Glu Val Phe Gly Asn Glu Ile Ser  *
                1700                1705

CCTGGGTGGG GCTGCTCCTC CAGGGCCACG TGCCAGGCCC CGGGCTGGCG GCTACTCAGC    1479

AGCCCTCCTC ACCCGTCTGG GGTTCAGCCC CTCCTCTGCC ACCTCCCCTA TCCACCCAGC    1539

CCATTCTCTC TCCTGTCCAA CCTAACCCCT TTCCTGCGGG CTTTTCCCCG GTCCCTTGAG    1599

ACCTCAGCCA TGAGGAGTTG CTGTTTGTTT GACAAAGAAA CCCAAGTGGG GGCAGAGGGC    1659

AGAGGCTGGA GGCAGGCCTT GCCCAGAGAT GCCTCCACCG CTGCCTAAGT GGCTGCTGAC    1719

TGATGTTGAG GGAACAGACA GGAGAAATGC ATCCATTCCT CAGGGACAGA GACACCTGCA    1779

CCTCCCCCCA CTGCAGGCCC CGCTTGTCCA GCGCCTAGTG GGGTCTCCCT CTCCTGCCTT    1839

ACTCACGATA AATAATCGGC CCACAGCTCC CACCCCACCC CCTTCAGTGC CCACCAACAT    1899

CCCATTGCCC TGGTTATATT CTCACGGGCA GTAGCTGTGG TGAGGTGGGT TTTCTTCCCA    1959

TCACTGGAGC ACCAGGCACG AACCCACCTG CTGAGAGACC CAAGGAGGAA AAACAGACAA    2019

AAACAGCCTC ACAGAAGAAT ATGACAGCTG TCCCTGTCAC CAAGCTCACA GTTCCTCGCC    2079

CTGGGTCTAA GGGGTTGGTT GAGGTGGAAG CCCTCCTTCC ACGGATCCAT GTAGCAGGAC    2139

TGAATTGTCC CCAGTTTGCA GAAAAGCACC TGCCGACCTC GTCCTCCCCC TGCCAGTGCC    2199

TTACCTCCTG CCCAGGAGAG CCAGCCCTCC CTGTCCTCCT CGGATCACCG AGAGTAGCCG    2259

AGAGCCTGCT CCCCCACCCC CTCCCCAGGG GAGAGGGTCT GGAGAAGCAG TGAGCCGCAT    2319

CTTCTCCATC TGGCAGGGTG GGATGGAGGA GAAGAATTTT CAGACCCCAG CGGCTGAGTC    2379

ATGATCTCCC TGCCGCCTCA ATGTGGTTGC AAGGCCGCTG TTCACCACAG GGCTAAGAGC    2439

TAGGCTGCCG CACCCCAGAG TGTGGGAAGG GAGAGCGGGG CAGTCTCGGG TGGCTAGTCA    2499
```

GAGAGAGTGT TTGGGGGTTC CGTGATGTAG GGTAAGGTGC CTTCTTATTC TCACTCCACC 2559

ACCCAAAAGT CAAAAGGTGC CTGTGAGGCA GGGGCGGAGT GATACAACTT CAAGTGCATG 2619

CTCTCTGCAG GTCGAGCCCA GCCCAGCTGG TGGGAAGCGT CTGTCCGTTT ACTCCAAGGT 2679

GGGTCTTTGT GAGAGTGAGC TGTAGGTGTG CGGGACCGGT ACAGAAAGGC GTTCTTCGAG 2739

GTGGATCACA GAGGCTTCTT CAGATCAATG CTTGAGTTTG GAATCGGCCG CATTCCCTGA 2799

GTCACCAGGA ATGTTAAAGT CAGTGGGAAC GTGACTGCCC CAACTCCTGG AAGCTGTGTC 2859

CTTGCACCTG CATCCGTAGT TCCCTGAAAA CCCAGAGAGG AATCAGACTT CACACTGCAA 2919

GAGCCTTGGT GTCCACCTGG CCCCATGTCT CTCAGAATTC TTCAGGTGGA AAAACATCTG 2979

AAAGCCACGT TCCTTACTGC AGAATAGCAT ATATATCGCT TAATCTTAAA TTTATTAGAT 3039

ATGAGTTGTT TTCAGACTCA GACTCCATTT GTATTATAGT CTAATATACA GGGTAGCAGG 3099

TACCACTGAT TTGGAGATAT TTATGGGGGG AGAACTTACA TTGTGAAACT TCTGTACATT 3159

AATTATTATT GCTGTTGTTA TTTTACAAGG GTCTAGGGAG AGACCCTTGT TTGATTTTAG 3219

CTGCAGAACT GTATTGGTCC AGCTTGCTCT TCAGTGGGAG AAAAACACTT GTAAGTTGCT 3279

AAACGAGTCA ATCCCCTCAT TCAGGAAAAC TGACAGAGGA GGGCGTGACT CACCCAAGCC 3339

ATATATAACT AGCTAGAAGT GGGCCAGGAC AGGCCGGGCG CGGTGGCTCA CGCCTGTAAT 3399

CCCAGCAGTT TGGGAGGTCG AGGTAGGTGG ATCACCTGAG GTCGGGAGTT CGAGACCAAC 3459

CTGACCAACA TGGAGAAACC CTGTCTCTAT TAAAAATACA AAAAAAAAAA AAAAAAAAAA 3519

TAGCCGGGCA TGGTGGCGCA AGCCTGTAAT CCCAGCTACT CAGGAGGCTG AGGCAGAAGA 3579

ATTGAACCCA GGAGGTGGAG GTTGCAGTGA GCTGAGATCG TGCCGTTACT CTCCAACCTG 3639

GACAACAAGA GCGAAACTCC GTCTTAGAAG TGGACCAGGA CAGGACCAGA TTTTGGAGTC 3699

ATGGTCCGGT GTCCTTTTCA CTACACCATG TTTGAGCTCA GACCCCCACT CTCATTCCCC 3759

AGGTGGCTGA CCCAGTCCCT GGGGGAAGCC CTGGATTTCA GAAAGAGCCA AGTCTGGATC 3819

TGGGACCCTT TCCTTCCTTC CCTGGCTTGT AACTCCACCA AGCCCATCAG AAGGAGAAGG 3879

AAGGAGACTC ACCTCTGCCT CAATGTGAAT CAGACCCTAC CCCACCACGA TGTGCCCTGG 3939

CTGCTGGGCT CTCCACCTCA GGCCTTGGAT AATGCTGTTG CCTCATCTAT AACATGCATT 3999

TGTCTTTGTA ATGTCACCAC CTTCCCAGCT CTCCCTCTGG CCCTGCTTCT TCGGGGAACT 4059

CCTGAAATAT CAGTTACTCA GCCCTGGGCC CCACCACCTA GGCCACTCCT CCAAAGGAAG 4119

TCTAGGAGCT GGGAGGAAAA GAAAAGAGGG GAAAATGAGT TTTTATGGGG CTGAACGGGG 4179

AGAAAAGGTC ATCATCGATT CTACTTTAGA ATGAGAGTGT GAAATAGACA TTTGTAAATG 4239

TAAAACTTTT AAGGTATATC ATTATAACTG AAGGAGAAGG TGCCCCAAAA TGCAAGATTT 4299

TCCACAAGAT TCCCAGAGAC AGGAAAATCC TCTGGCTGGC TAACTGGAAG CATGTAGGAG 4359

AATCCAAGCG AGGTCAACAG AGAAGGCAGG AATGTGTGGC AGATTTAGTG AAAGCTAGAG 4419

ATATGGCAGC GAAAGGATGT AAACAGTGCC TGCTGAATGA TTTCCAAAGA GAAAAAAAGT 4479

TTGCCAGAAG TTTGTCAAGT CAACCAATGT AGAAAGCTTT GCTTATGGTA ATAAAAATGG 4539

CTCATACTTA TATAGCACTT ACTTTGTTTG CAAGTACTGC TGTAAATAAA TGCTTTATGC 4599

AAACC 4604

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 428 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Met Glu Ala Met Ala Ala Ser Thr Ser Leu Pro Asp Pro Gly Asp Phe
1 5 10 15

Asp Arg Asn Val Pro Arg Ile Cys Gly Val Cys Gly Asp Arg Ala Thr
20 25 30

Gly Phe His Phe Asn Ala Met Thr Cys Glu Gly Cys Lys Gly Phe Phe
35 40 45

Arg Arg Ser Met Lys Arg Lys Ala Leu Phe Thr Cys Pro Phe Asn Gly
50 55 60

Asp Cys Arg Ile Thr Lys Asp Asn Arg Arg His Cys Gln Ala Cys Arg
65 70 75 80

Leu Lys Arg Cys Val Asp Ile Gly Met Met Lys Glu Phe Ile Leu Thr
85 90 95

Asp Glu Glu Val Gln Arg Lys Arg Glu Met Ile Leu Lys Arg Lys Glu
100 105 110

Glu Glu Ala Leu Lys Asp Ser Leu Arg Pro Lys Leu Ser Glu Glu Gln
115 120 125

Gln Arg Ile Ile Ala Ile Leu Leu Asp Ala His His Lys Thr Tyr Asp
130 135 140

```
Pro Thr Tyr Ser Asp Phe Cys Gln Phe Arg Pro Pro Val Arg Val Asn
145                 150                 155                 160

Asp Gly Gly Gly Ser His Pro Ser Arg Pro Asn Ser Arg His Thr Pro
                165                 170                 175

Ser Phe Ser Gly Asp Ser Ser Ser Ser Cys Ser Asp His Cys Ile Thr
            180                 185                 190

Ser Ser Asp Met Met Asp Ser Ser Ser Phe Ser Asn Leu Asp Leu Ser
        195                 200                 205

Glu Glu Asp Ser Asp Asp Pro Ser Val Thr Leu Glu Leu Ser Gln Leu
    210                 215                 220

Ser Met Leu Pro His Leu Ala Asp Leu Val Ser Tyr Ser Ile Gln Lys
225                 230                 235                 240

Val Ile Gly Phe Ala Lys Met Ile Pro Gly Phe Arg Asp Leu Thr Ser
                245                 250                 255

Glu Asp Gln Ile Val Leu Leu Lys Ser Ser Ala Ile Glu Val Ile Met
            260                 265                 270

Leu Arg Ser Asn Glu Ser Phe Thr Met Asp Asp Met Ser Trp Thr Cys
        275                 280                 285

Gly Asn Gln Asp Tyr Lys Tyr Arg Val Ser Asp Val Thr Lys Ala Gly
    290                 295                 300

His Ser Leu Glu Leu Ile Glu Pro Leu Ile Lys Phe Gln Val Gly Leu
305                 310                 315                 320

Lys Lys Leu Asn Leu His Glu Glu Glu His Val Leu Leu Met Ala Ile
                325                 330                 335

Cys Ile Val Ser Pro Asp Arg Pro Gly Val Gln Asp Ala Ala Leu Ile
            340                 345                 350

Glu Ala Ile Gln Asp Arg Leu Ser Asn Thr Leu Gln Thr Tyr Ile Arg
        355                 360                 365

Cys Arg His Pro Pro Pro Gly Ser His Leu Leu Tyr Ala Lys Met Ile
    370                 375                 380

Gln Lys Leu Ala Asp Leu Arg Ser Leu Asn Glu Glu His Ser Lys Gln
385                 390                 395                 400

Tyr Arg Cys Leu Ser Phe Gln Pro Glu Cys Ser Met Lys Leu Thr Pro
                405                 410                 415

Leu Val Leu Glu Val Phe Gly Asn Glu Ile Ser *
            420                 425
```

What is claimed is:

1. A method comprising, identifying individuals having a certain phenotype, determining the presence or absence of gene markers associated with the phenotype, and instituting a lifestyle change to exploit or counteract the phenotype expressed by the gene marker.

2. The method of claim 1 wherein the phenotype is low level HDL-C and HDL_2-C, the gene marker is APO E2 and exercise training is instituted to improve the HDL-C and HDL_2-C levels.

3. The method of claim 1 wherein the phenotype is hypertension, the gene marker is at least one insertion ("I") ACE allele and exercise training is instituted to decrease systolic and diastolic blood pressure.

4. The method of claim 1 wherein the phenotype is a reduction of bone mineral density, the gene marker is BB vitamin D receptor and exercise training is instituted to increase bone mineral density.

5. The method of claim 1 wherein the phenotype is a determined VO_2max value for a determined habitual level of physical activity, the gene marker is at the angiotensinconverting enzyme locus and a specific aerobic sport is conducted to exploit, for sport, the presence of the marker.

6. A method of determining the risk of cardiovascular disease in an individual comprising, determining the ACE phenotype of the individual.

7. The method of claim 1 wherein the phenotype is hypertension, the gene marker is

located at restriction cites of the lipoprotein lipase gene locus and exercise training is instituted to decrease blood pressure.

8. The method of claim 7 wherein the gene marker is at the LPL PvuII restriction site.

9. The method of claim 7 wherein the gene marker is at the HindIII restriction site.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/22974

A. CLASSIFICATION OF SUBJECT MATTER
IPC(6) :C12Q 1/68
US CL : 436/6
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 436/6; 536/24.31

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	HAGBERG et al. Does exercise training play a role in the treatment of essential hypertension? Journal of Cardiovascular Risk. August 1995, Vol. 2, No. 4, pages 296-302, see entire document.	1-3, 5-9
X ---, P Y	SAKAI et al. Angiotensin-Converting-Enzyme Gene Polymorphism predicts the Depressor Effect of Exercise Therapy in Hypertensives. Journal of the American College of Cardiology. February 1997, Vol. 29, No 2(Suppl A) page 84A, see entire abstract 724-5.	1,3 --- 5,6

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 25 MARCH 1998	Date of mailing of the international search report 05 MAY 1998
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	Authorized officer DEBRA SHOEMAKER

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/22974

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X -- Y	RAURAMAA et al. Effect of Exercise Training on Lipoprotein Cholesterol with Special Reference to Apolipoprotein E (ApoE) Polymorfism. Circulation. October 1991, Vol. 84, No. 4(Suppl. 2), page II-119, see entire abstract 0475.	1 -- 2
X	PRATLEY et al. Insulin Resistance, Hyperinsulinemia, and Increased Sympathetic Nervous System Activity Associated with Hypertension Improve with Diet and Exercise. Diabetes. May 1994, Vol. 43, Suppl. 1, page 46A, see entire abstract 143.	1
X	BENLIAN et al. Premature Atherosclerosis in Patients with Familial Chylomicronemia Caused by Mutations in the Lipoprotein Lipase Gene. New England Journal of Medicine. September 1996, Vol. 335, No. 12, pages 848-854, especially abstract and page 848.	1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/22974

B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

Searched inventors and keywords: exercise or aerobic and hypertension, cholesterol, cardiovascular disease or bone density and apoE or LPL, or VDR or ACE in APS, CAPLUS, MEDLINE, SCISEARCH BIOSIS, WPIDS.

THIS PAGE BLANK (USPTO)